

Result No.	Score	Query Match Length	DB ID	Description
--	--	--	--	Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.
		%		SUMMARIES
				Result No. Score Query Match Length DB ID Description
1	164.6	90.7	321	AO21520 Human drug metabol
2	164.6	90.7	340	23 AAB14442 Human drug metabol
3	93.5	51.5	175	23 ABB04738 Human heparan sulf
4	78.0	43.0	307	20 AAY17063 Human 3-OST-1 prot
5	78.0	43.0	307	24 AAO26506 Human 3-OST-1 prot
6	78.0	43.0	321	23 ABP4127 Human ovarian anti
7	772.5	42.6	255	23 ABO21521 255mer human drug
8	75.8	41.8	307	22 ABB30928 Peptide #3579 enco
9	75.8	41.8	307	22 ABB36116 Peptide #3622 enco

FT	Region	161..163 /note= "Protein kinase C phosphorylation site"	Db	191 DPNTCEVNTKYKAVRTSIYTQKHLERMLKYPPLQFFVVGDRLITPELPELOVKEPLNL 250
FT	Modified-site	179..182 /note= "Asn is N-glycosylated"	Qy	276 PERISQONLYFATGTYCILRKENIIIPKCLAGSKGRHPEVDPSVTKLREFHPPNQKF 335
FT	Modified-site	262..265 /note= "Asn is N-glycosylated"	Db	251 PERISQONLYFATGTYCILRKENIIIPKCLAGSKGRHPEVDPSVTKLREFHPPNQKF 310
FT	Region	314..316 /note= "Protein kinase C phosphorylation site"	Qy	336 YQITGTLINWP 346
FT	XX	W0200242437-A2.	Db	311 YQITGTLINWP 321
PN	XX			
PD	XX	30-MAY-2002.	RESULT 2	
PF	XX	16-NOV-2001; 2001WO-US43109.	AAE14442	
PR	XX	27-NOV-2000; 2000US-252895P.	ID : AAE14442 standard; Protein; 340 AA.	
PR	XX	14-DEC-2000; 2000US-0735935.	AC	
PA	XX	(PEKE ) PE CORP NY.	XX	
PI	XX	Guegler K, Ketchum RA, Di Francesco V, Beasley EM;	XX	Human drug metabolising enzyme (DME) -5.
DR	XX	WPI: 2002-479949/51.	KW	Human; drug metabolising enzyme; DME-5; autoimmune; inflammatory;
N-PSDB	XX	DR: AAL28608, AAL28608.	CC	cell; proliferative; developmental; endocrine; eye; metabolic; AIDS;
XX			CC	gastrointestinal disorder; liver disorder; cancer; arteriosclerosis;
PT		New human drug-metabolizing enzyme, related to the sulfotransferase	CC	adult respiratory distress syndrome; anaemia; epilepsy; hypothyroidism;
PT		PT drug-metabolizing enzyme subfamily, useful as a model for developing	CC	hypothalamus; pituitary; diabetes; hypogonadism; conjunctivitis;
PT		human therapeutic targets and as a target for human therapeutics -	CC	glaucoma; cystic fibrosis; hypercholesterolaemia; gastritis;
PS	XX	Claim 1; Fig 2; 65pp; English.	CC	peptic ulcer; hepatitis; gene therapy.
XX	OS		XX	Homo sapiens.
XX	OS		XX	PN WO200190334-A2.
XX	OS		XX	PN PN 29-NOV-2001.
CC	CC	The invention relates to an isolated human drug-metabolising polypeptide	XX	PD PD 29-NOV-2001.
CC	CC	which is related to the sulfotransferase drug-metabolising enzyme	XX	XX 25-MAY-2001; 20001WO-US17150.
CC	CC	subfamily, consisting or comprising a sequence of 321 amino acids,	XX	XX 25-MAY-2000; 20000US-207901P.
CC	CC	given in the specification, or its fragment comprising 10 contiguous	PR PR 01-JUN-2000; 20000US-20893P.	
CC	CC	amino acids, or an amino acid sequence of an allelic variant or ortholog	PR PR 07-JUN-2000; 20000US-20983P.	
CC	CC	The protein of the invention, its allelic variant or ortholog or	PR PR 15-JUN-2000; 20000US-211935P.	
CC	CC	fragment, is used to identify a modulator of a human drug-metabolising	PR PR 22-JUN-2000; 20000US-213744P.	
CC	CC	enzyme protein. The drug-metabolising enzyme proteins isolated from	XX XX PA (INCY-) INCYTE GENOMICS INC.	
CC	CC	humans and their human/mammalian orthologues serve as targets for	XX XX Yue H, Sanjanwala MS, Baughn MR, Gandhi AR, Ring HZ, Elliott V;	
CC	CC	identifying agents for use in mammalian therapeutic applications, and	PI Walia NK, Yang J, Khan FA, Ramkumar J, Tang YT, Hafalia A, Lal P;	
CC	CC	biological assays related to the drug-metabolising enzyme proteins. The	PI Nguyen DB, Yao MG, Lee EA, Troubley CM, Patterson C, Lu Y;	
CC	CC	proteins can also be used in screening assays to screen a compound for	PI Burford N, Ding L, Bruns CM, Kearney L, Reddy R;	
CC	CC	its ability to stimulate or inhibit interaction between drug-metabolising	XX DR WPI; 2002-09765/13.	
CC	CC	enzymes and a molecule that normally interacts with the drug-	DR DR N-PSDB; AAP24010.	
CC	CC	metabolising enzyme proteins. The proteins also provide a target for	XX PT New human drug metabolizing enzymes and polynucleotides encoding the	
CC	CC	diagnosing a disease or predisposition to disease mediated by the	PT enzyme for diagnosing, preventing or treating cell proliferative,	
CC	CC	peptide, and in pharmacogenomic analysis. The nucleic acid encoding the	PT autoimmune/inflammatory, endocrine, eye, metabolic and gastrointestinal	
CC	CC	protein of the invention also provides vectors for gene therapy in	PT disorders -	
CC	CC	patients with aberrant expression of a gene encoding the drug-	XX Claim 1; Page 140-141; 158pp; English.	
CC	CC	metabolising enzyme. This sequence represents the human drug metabolising	XX PT	
XX	Sequence	321 AA;	XX	XX
Qy	Best Local Similarity	90.7%; Score 1646; DB 23; Length 321;	CC	The present sequence is human drug metabolising enzyme (DME) -5.
Matches	311; Conservative	0; Mismatches 0; Indels 0; Gaps 0;	CC	DME polypeptide, polynucleotide and modulators are useful for
Qy	36	RLOPICPTEGLGGARTAQEFPLRALQFRQLLHEFRKGNAKSKQVRLHDLYQQLPKAI 95	CC	diagnosis, treatment and prevention of autoimmune/inflammatory,
Db	11	RLOPICPTEGLGGARTAQEFPLRALQFRQLLHEFRKGNAKSKQVRLHDLYQQLPKAI 70	CC	cell; proliferative, developmental, endocrine, eye, metabolic,
Qy	96	IGVRKGGSFTRALLEMLNLHFAVVAQSDFHFDENTGKGIWYRKMPFSYSPQQTIEK 155	CC	CC and gastrointestinal disorders, including liver disorders.
Db	71	IGVRKGGSFTRALLEMLNLHFAVVAQSDFHFDENTGKGIWYRKMPFSYSPQQTIEK 130	CC	The autoimmune/inflammatory disorders treatable include
Qy	156	SPAYFITEEPVERIYKONSIKLLIVREPTRAISDTQVLEGKERIKNTYKFERLAI 215	CC	AIDS, adult respiratory distress syndrome, Addison's disease,
Db	131	SPAYFITEEPVERIYKONSIKLLIVREPTRAISDTQVLEGKERIKNTYKFERLAI 190	CC	allergies, anaemia, asthma, atherosclerosis, osteoporosis, autoimmune
Qy	216	DPNTCEVNTKYKAVRTSIYTQKHLERMLKYPPLQFFVVGDRLITPELPELOVKEPLNL 275	CC	haemolytic anaemia, autoimmune thyroiditis, Crohn's disease, atopic
			CC	dermatitis, diabetic mellitus, Graves' disease, glomerulonephritis,
			CC	rheumatoid arthritis, scleroderma, systemic lupus erythematosus,
			CC	systemic sclerosis, ulcerative colitis, haemodialysis and ureitis,
			CC	viral, bacterial, fungal, parasitic, protozoal, helminthic infections



XX	WPI; 1999-312968/26.
DR	N-PSDB; AAX3746.
XX	Reparan sulfate D-glucosaminyl 3-O-sulfotransferase and related polynucleotides
XX	Claim 17; Page 74-75; 95pp; English.
CC	The invention relates to nucleic acid molecules (AAX37245-X37250) encoding heparan sulfate D-glucosaminyl 3-O-sulfotransferases (3-OSTs). The 3-OST proteins can be used for 3-O-sulfating saccharide residues within a preparation of glycosaminoglycan or proteoglycan polysaccharides. 3-OST-1 can be used for enriching the antithrombin-binding fraction in a preparation of heparan sulfates (HS). 3-OST-1 can also convert HS proteoglycan anticoagulant precursor to the HS anticoagulant (HSact). The 3-OSTs (optionally lacking enzymatic function) can be used to determine partial sequence information for complex polysaccharides. The 3-OST proteins, genes and antibodies are also useful for diagnosis of disorders involving HS biosynthesis. 3-OSTs are especially used to generate anticoagulant pentasaccharides, which may be used to treat thrombotic disorders such as deep vein thrombosis and pulmonary embolism. Coagulant enzyme inactivation by antithrombin is enhanced by complexing of antithrombin with endothelial cell surface HS proteoglycans. This is responsible for the non-thrombogenic properties of blood vessels. The present sequence represents a human 3-OST-1.
SQ	Sequence 307 AA;
Query Match	42.0%; Score 780; DB 20; Length 307;
Best Local Similarity	46.4%; Pred. No. 4.5e-69;
Matches	153; Conservative 64; Mismatches 89; Indels 24; Gaps 6;
Qy	16 LGSIAVGSILYLVARVGSDLRLQICPIEGRGGARTAQEFPLRALQFGRGLLHEFRKGK 75
Db	1 MAALLGAVL-LVAGQ----PQLVSRP-----PELGQCBLLRAGTLQD---- 39
DY	76 ASKEQVRVLHDLYQOLPKAALLIGYRKKGTRALLEMLNLIHAVVRSQEIEFFDENDYNGK 135
Db	40 DVRDGVAPNGSAQQLPQTLIGYRKKGTRALLEMLSLHDVAALENEFFWEHYSHG 99
DY	136 IEWFRKMQPSYPOQITPSAFVTFEPYKMNNSKIKLIVTPPTTAISDTQ 195
Db	100 LGWYLSQMEFSWPHQLTVEKTPAYFTSPVPERYSMSNIRLILRPSERVLSDTQ 159
DY	196 VLEGERKRNKTYKFEKLAIIDPNTECVNPKYKAVRTSITTKHLERWLXKPIEOFHVDG 255
Db	160 VFTNHMQKRPYPSLVEEFLVRDG--RLNDYKALNRSLYTHMOMNLRFPLSHIHTVDG 217
DY	256 DRLETEPLELQLVEKELNLPPRISQVNLYFNATRGFCYLFRTNTIFNCCLAGSKGRHPE 315
Db	218 DRLLRDPEFLPKYERFLPSQINASNFNKTKGFCYL-DSGRDRLHESKGRAFPQ 276
DY	316 VDPBPVITKRRFFFPLFNQFQYQTGRTNW 345
Db	277 VDPKLNKCHHEYFEPNKKFELVGRTFDW 306
RESULT 5	
IID	A026606 standard; Protein; 307 AA.
XX	A026606;
XX	28-MAR-2003 (first entry)
XX	Human 3-OST-1 protein sequence.
XX	Anticoagulant; ahrombolytic sulfate; 6-O position; N-acetylglucosamine; GlcNAc; sugar residue; glucosaminyl 6-O-sulfotransferase; 6-Ost protein; heparan sulfate; 3-O-sulfated; 3-OST-1 protein; antithrombin; extracorporeal medical device; dialysis machine; thrombotic disease; extracorporeal implant; intracorporeal device; transplant; stent; implant.
XX	136 IENYRKMPFESPOQITPSAFVTFEPYKMNNSKIKLIVTPPTTAISDTQ 195
Db	1 MAALLGAVL-LVAGQ----PQLVSRP-----AELGQCBLLRAGTLQD---- 39
Qy	76 ASKEQVRVLHDLYQOLPKAALLIGYRKKGTRALLEMLNLIHAVVRSQEIEFFDENDYNGK 135
Db	40 DVRDGVAPNGSAQQLPQTLIGYRKKGTRALLEMLSLHDVAALENEFFWEHYSHG 99
Qy	16 LGSIAVGSILYLVARVGSDLRLQICPIEGRGGARTAQEFPLRALQFGRGLLHEFRKGK 75
Db	1 MAALLGAVL-LVAGQ----PQLVSRP-----AELGQCBLLRAGTLQD---- 39
Qy	277 VDPKLNKCHHEYFEPNKKFELVGRTFDW 306
XX	28-MAR-2003 (first entry)

Qy	256 DRLLTEPLDELQVEKFLNLPPIRISOTVNLVFNATRGYCLRFNTIIFRNKLAGSKGRTHPE	315	CC polynucleotides may also be used in screening for compounds which
Db	218 DRLLTEPLDELQVEKFLNLPPIRISOTVNLVFNATRGYCLRFNTIIFRNKLAGSKGRTHPE	276	CC modulate ovarian antigen expression or activity. The polynucleotides may further be used for gene therapy, chromosome mapping, in the
Qy	316 VDPSSVITKLKKFPHPPENOKFYQITGRTLNW	345	CC identification of individuals and in forensic analysis, and the
Db	277 VDPKLNNKLUHEYFHEPNKRFELVGRTFW	306	CC polypeptides may be used as food additives or to prepare antibodies useful in disease diagnosis, drug targeting and phenotyping. The present
RESULT 6			CC sequence represents a human ovarian antigen of the invention.
ABP41427			CC Note: The sequence data for this patent did not form part of the printed
ID ABP41427 standard; Protein: 321 AA.			CC specification, but was obtained in electronic format directly from WIPO
XX			CC at ftp.wipo.int/pub/published_pct_sequences.
AC			XX
XX	Sequence 321 AA:		XX
DT	Query Match 43.0%; Score 780; DB 23; Length 321;		
XX	Best Local Similarity 46.4%; Pred. No. 4.8e-89;		
DE	Mismatches 89; Indels 24; Gaps 6;		
XX	Matches 153; Conservative 64; Mimatches 89;		
DE	CC		
Qy	16 LGSLAVSLLIVARYGSLDLQIPICTIEGLGGARTQAEFLPLRAOKRGLLHEFRKGN	75	CC
Db	15 MAALLIGAVL-LVAQ-----PQLVPSRP-----AELGQQELRKAGTLQD-----	53	CC
Qy	76 ASKEQYRLHDIVQQLPAATTIGVKGCTTRALLEMLNHAVVTKASDTHFFNDENDYKKG	135	CC
Db	54 DVRDGAPNGSAQQLQTIIGVVKGCTTRALLEMLSHPDVAAANEVHFFDWEENHSHG	113	CC
Qy	136 IEWYRKMPFESTPQOIIIEKSPAYFETTEEVPERITYKDNNSKIKLILIVREPTPRAISDYTQ	195	CC
Db	114 LGWLYSMQMPFSWPHQIIVKTPAYFSPKVPERVSNPSTRLLLRDPSRVLSDYTQ	173	CC
Qy	196 VLEGKSRKNCYKYYKEEKLAIADPNTCEVNTKYKAVRTSTIYTKHLERNLKYKFPFBQFHVVDG	255	CC
Db	174 VFNHNGQHKCYSPTSEBFVLVDG--RLNVYDKALNRSLYHVMONVLFREFPLRHILVVDG	231	CC
Qy	256 DRLLTEPLDELQVEKFLNLPPIRISOTVNLVFNATRGYCLRFNTIIFRNKLAGSKGRTHPE	315	CC
Db	232 DRLLTEPLDELQVEKFLNLPPIRISOTVNLVFNATRGYCLRFNTIIFRNKLAGSKGRTHPE	290	CC
Qy	316 VDPSSVITKLKKFPHPPENOKFYQITGRTLNW	345	CC
Db	291 VDPKLNNKLUHEYFHEPNKRFELVGRTFW	320	CC
RESULT 7			
AAO21521			XX
ID	AAO21521 standard; Protein: 255 AA.		XX
XX			XX
AC			XX
XX			XX
DT			XX
16-AUG-2002	(first entry)		XX
255-mer human drug metabolising enzyme related protein.			
Homo sapiens.			XX
XX			XX
PN			XX
W0200200677-A1.			XX
XX			XX
PD			XX
03-JAN-2002.			XX
XX			XX
PP			XX
07-JUN-2001;	2001WO-US18569.		XX
XX			XX
PR			XX
07-JUN-2000;	2000US-209467P.		XX
XX			XX
PA			XX
(HUMA-) HUMAN GENOME SCI INC.			XX
XX			XX
PI			XX
Birse CE,	Rosen CA;		XX
XX			XX
WPI: 2002-147878/19.			XX
DR			XX
N-PSDB;	ABQ54504.		XX
XX			XX
Isolated nucleic acid molecules encoding novel ovarian polypeptides, and also			XX
useful in the prevention, treatment and diagnosis of cancer (e.g. ovarian cancer), immune disorders, cardiovascular disorders and neurological diseases -			XX
XX			XX
Claim 11; SEQ ID No 2559; 2922pp; English.			OS
XX			PN
The invention relates to 2175 novel human ovarian antigens (ABP41054-ABP4228) and to cDNAs encoding them (ABQ54305), and also			XX
CC encompasses polypeptides 90% identical and polynucleotides 95% identical			XX
CC to the sequences of the invention. The invention additionally relates to recombinant vectors and host cells comprising human ovarian antigen polynucleotides, antibodies against human ovarian antigens, and the use			XX
CC of ovarian antigen polynucleotides and polypeptides in diagnosing, treating, prognosing or preventing various ovary and/or breast-related			XX
CC disorders. Such conditions include ovarian cancer and breast cancer, and			XX
CC metastatic tumours of ovarian or breast origin, reproductive system			XX
CC disorders (e.g. infertility, disorders of pregnancy, anovulation,			XX
CC polycystic ovary syndrome, ovarian cysts, and dysmenorrhoea), endocrine			XX
CC disorders, infections (e.g., chlamydia, HIV, toxoplasmosis, and toxic			XX
CC shock syndrome), inflammatory conditions (e.g., mastitis, cophritis and			XX
CC vaginitis), immune disorders (e.g., congenital and acquired			XX
CC immunodeficiencies, autoimmune ophoritis, systemic lupus erythematosus),			XX
CC blood-related disorders (e.g., anaemia), cardiovascular disorders,			XX
CC respiratory disorders, neurological disorders, gastrointestinal disorders			XX
CC and urinary system disorders. Ovarian antigen polypeptides and			XX
XX			PT
Guegler K, Ketchem RA, Di Francesco V, Beasley EM;			PI
XX			XX
WPI; 2002-479949/51.			DR
XX			XX
New human drug-metabolizing enzyme, related to the sulfotransferase			PT

drug-metabolizing enzyme subfamily, useful as a model for developing human therapeutic targets and as a target for human therapeutics -

XX Disclosure: Fig 2; 65pp; English.

CC The invention relates to an isolated human drug-metabolising polypeptide subfamily, consisting or comprising of a sequence of 321 amino acids, given in the specification, or its fragment comprising 10 contiguous amino acids, or an amino acid sequence of an allelic variant or ortholog. The protein of the invention, its allelic variant or ortholog or fragment, is used to identify a modulator of a human drug-metabolising enzyme protein. The drug-metabolising enzyme proteins isolated from humans and their human/mammalian orthologues serve as targets for identifying agents for use in mammalian therapeutic applications, and biological assays related to the drug-metabolising enzyme proteins. The proteins can also be used in screening assays to screen a compound for its ability to stimulate or inhibit interaction between drug-metabolising enzymes and a molecule that normally interacts with the drug-metabolising enzyme proteins. The proteins also provide a target for diagnosing a disease or predisposition to disease mediated by the peptide, and in pharmacogenomic analysis. The nucleic acid encoding the protein of the invention also provides vectors for gene therapy in patients with aberrant expression of a gene encoding the drug-metabolising enzyme. This sequence represents a 255-mer human protein related to the human drug metabolising enzyme protein of the invention.

XX Sequence 255 AA;

Query Match 42.6%; Score 772.5; DB 23; Length 255;  
Best Local Similarity 53.1%; Pred. No. 1.9e-68; Mismatches 52; Indels 3; Gaps 2;  
Matches 137; Conservative 52; Mismatches 66; Indels 3; Gaps 2;  
Qy 88 QDPLRATIGGVRKGTRALLEMLNLPAYVKAQSBIHFFDNDENYGRKGTEWYKRMDFSY 147  
Db 1 QQLPQTIIIGGVRKGTRALLEMLSLHPDVAAENEVHFDWEEHYSHGGWILSQMPFW 60  
Qy 148 PQQTIIEKSPAYFITEEVPERYKONSSKLILLIVREPTTRASDVTQLEGERKNTY 207  
Db 61 PHQUTVKEPKPAYFSPKVPERVSNPSSLRLLIRDPEVRLSDTYQFYNHMOKKPY 120  
Qy 208 YKPEKLAIDPNTCVNTKXKAVRTSITYKHLERWKYFIEQHVVDGDLTTEPLPLQ 267  
Db 121 PSKEEFLVRDG--RLNVDTKALNRSLYHFMONWLRFPLRHITHIVDGRLLRDPEFEIQ 178  
Qy 268 LVEKPLNLPRISQNLNYNATRGFYCLRFENLIFNKCLAGSKGRHFDPSVTKLRFK 327  
Db 179 KVERFLKLSPQINASNPFKTRGFCYL--DSGDRCLHESESGRAPQDPKLNKHEY 237  
Qy 328 FHEPNOKRYQITGRTLNW 345  
Db 238 FHEPNKKEFELVGRTFDQ 255

RESULT 8  
ID ABB30928 standard; Peptide; 307 AA.  
XX ABB30928:  
XX AC:  
XX DT 01-FEB-2002 (first entry)  
XX Peptide #3579 encoded by breast cell single exon nucleic acid probe.  
XX Human; microarray; single exon probe; gene expression; breast;  
XX disease; cancer.  
XX Homo sapiens.  
XX WO2001137271-A2.  
XX PD 09-AUG-2001.  
XX

Query Match 41.8%; Score 758; DB 22; Length 307;  
Best Local Similarity 46.2%; Pred. No. 7e-67; Mismatches 63; Indels 24; Gaps 6;  
Matches 150; Conservative 63; Mismatches 88; Indels 24; Gaps 6;  
Qy 16 IGSLSLAGSLLIVARYGSLDRLQPICTIEGRGGARTQAEFFPLAQFKRGLLHEPRKGN 75  
Db 7 MEALLLAVL-LVAQ---PQLVPSP-----AEUQQEELLRKAGPLQD---- 45  
Qy 76 ASKEQVRLHDLVQQLPKAIIGGVRKGTRALLEMLNLPAYVKAQSBIHFFDNDENYGRKG 135  
Db 46 DVRDGAPNGSAQQLSotigIIVGRKGTRALLEMLSLHPDVAAENEVHFDWEEHYSHG 105  
Qy 136 IEWIKRKKPFSQQTLERSPAYFITEEVPERYKONSSKLILLIVREPTTRASDVTQ 195  
Db 106 LGWLYLSQMPFSWPHQLTVEKTPAYFSPKVPERVSNPSSLRLLIRDPSVLSDTQ 165  
Qy 196 VLEGKRRKRNKTYKPFKLAIDNTCEVNTKXKAVRTSITYKHLERWKYFIEQHVVDG 255  
Db 166 VEPYRNMOKHKKYKPPSIEFLVGD--FLNVTKALNRSLYHFMONWLRFPLRHITHIVD 223  
Qy 256 DRLLTEPLPELQVKEPLNLPRISQNLNYNATRGFYCLRFENLIFNKCLAGSKGRHPE 315  
Db 224 DRLLRDPFPEIQVERFLKLSPQINASNPFKTRGFCYL--DSGDRCLHESESGRAPQ 282  
Qy 316 VDPSVTKLKREFHPFNQKFQITG 340  
Db 283 VDPKLUKLHBYFHEPNKCFELVG 307

RESULT 9  
 ABB36116 standard; Peptide; 307 AA.  
 XX  
 AC  
 XX  
 DT 04-FEB-2002 (first entry)  
 XX  
 DB Peptide #3622 encoded by human foetal liver single exon probe.  
 XX Human; foetal liver; gene expression; single exon nucleic acid probe.  
 XX Homo sapiens.  
 OS XX  
 PN WO200157277-A2.  
 XX 09-AUG-2001.  
 PD XX  
 PP 30-JAN-2001; 2001WO-US00669.  
 PR XX  
 PR 04-FEB-2000; 2000US-0180312.  
 PR 26-MAY-2000; 2000US-0207456.  
 PR 30-JUN-2000; 2000US-0608408.  
 PR 03-AUG-2000; 2000US-0632366.  
 PR 21-SEP-2000; 2000US-0234687.  
 PR 27-SEP-2000; 2000US-0236359.  
 PR 04-OCT-2000; 2000GB-0024263.  
 XX (MOLE-) MOLECULAR DYNAMICS INC.  
 PA XX  
 PI Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX WPI; 2001-483447/52.  
 PT Human genome-derived single exon nucleic acid probes useful for  
 PT analyzing gene expression in human fetal liver -  
 XX PS Claim 27; SEQ ID NO 28751; 639pp + sequence listing; English.  
 CC The invention relates to a single exon nucleic acid probe for  
 CC measuring human gene expression in a sample derived from human foetal  
 CC liver. The single exon nucleic acid probes may be used for predicting,  
 CC measuring and displaying gene expression in samples derived from human  
 CC fetal liver. The present sequence is a peptide encoded by a single exon  
 CC nucleic acid probe of the invention.  
 CC Note: The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WIPO at [ftp.wipo.int/pub/published\\_pct\\_sequences](http://wipo.int/pub/published_pct_sequences).  
 XX Sequence 307 AA;  
 SQ Query Match 41.8%; Score 758; DB 22; Length 307;  
 Best Local Similarity 46.2%; Pred. No. 7e-67; Mismatches 88; Indels 24; Gaps 6;  
 Matches 150; Conservative 63; N mismatches 88;  
 Db 16 LGSIAVSSLVLYARVGSLLDRLQIPIGTRGGARTQAEPIRALQFRKGN 75  
 7 MALLGAVL-LVAQ---POLYPSRP-----ALGQQELLRKAGTLQD---- 45  
 Qy 17 ASKEQVRHLHDYQOLPKAIILIGYRKGETRALLEMLNLHPAVTKASQEIHFDDNDENYKGK 135  
 46 DVRDGAEPNGSACQLPQTIIIGYRKGETRALLEMLSLHPDVAAENHFFDWEHYSG 105  
 SQ 136 IEWYRKGMFSPSQQITKSPAYFTEEVPERIYKQNSKICLILIVTPPTRAIDYQ 195  
 Db 106 LGNYLSQMFPSWHQLVATEPKPAYTSKPVYVSNPSTRILLLRDPSRVLSDYQ 165  
 Qy 196 VLEGKERKNTKYKFEKLAIDPNTCEVNTKYKAVRTSLYTKHLERWTFPLQEFHVVG 255  
 Db 166 VYNHMQKHPKPSIEFLVRDG--RLNVDYKALNRSLYHVMQNWRFPFLRHIVDG 223  
 Qy 256 DRLIETPLPELOVKEFLNLPPIRSOYNLYENATRGFCYLRENNIFKCLAGSKGRTHPE 315

RESULT 10  
 ABB21504 standard; Protein; 307 AA.  
 XX ID ABB21504  
 AC XX  
 AC ABB21504;  
 XX DT 23-JAN-2002 (first entry)  
 DE Protein #3503 encoded by probe for measuring heart cell gene expression.  
 XX Human; gene expression; heart; microarray; vascular system;  
 KW cardiovascular disease; hypertension; cardiac arrhythmia;  
 KW congenital heart disease.  
 XX OS Homo sapiens.  
 XX PN WO200157274-A2.  
 XX PD 09-AUG-2001.  
 XX PR 30-JAN-2001; 2001WO-US00666.  
 XX PR 04-FEB-2000; 2000US-018012.  
 PR 26-MAY-2000; 2000US-0207456.  
 PR 30-JUN-2000; 2000US-0608408.  
 PR 03-AUG-2000; 2000US-0632366.  
 PR 21-SEP-2000; 2000US-0234687.  
 PR 27-SEP-2000; 2000US-0236359.  
 PR 04-OCT-2000; 2000GB-0024263.  
 XX PA (MOLB-) MOLECULAR DYNAMICS INC.  
 XX PI Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX DR WPI; 2001-488899/53.  
 XX PR 03-AUG-2000; 2000US-0632366.  
 PR 21-SEP-2000; 2000US-0234687.  
 PR 27-SEP-2000; 2000US-0236359.  
 PR 04-OCT-2000; 2000GB-0024263.  
 XX PS Claim 15; SEQ ID NO 23274; 530pp; English.  
 XX CC The present invention relates to single exon nucleic acid probes for  
 CC measuring human gene expression in a sample derived from human heart (see  
 CC ABA21535-ABA41305). The present sequence is a protein encoded by one such  
 CC probe. The probes may be used for predicting, measuring and displaying  
 CC gene expression in samples derived from the human heart via microarrays.  
 CC By measuring gene expression, the probes are useful for predicting,  
 CC diagnosing, grading, staging, monitoring and prognosis diseases of the  
 CC human heart and vascular system, e.g., cardiovascular disease.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at [ftp.wipo.int/pub/published\\_pct\\_sequences](http://wipo.int/pub/published_pct_sequences).  
 XX Sequence 307 AA;  
 SQ Query Match 41.8%; Score 758; DB 22; Length 307;  
 Best Local Similarity 46.2%; Pred. No. 7e-67; Mismatches 88; Indels 24; Gaps 6;  
 Matches 150; Conservative 63; N mismatches 88;  
 Db 16 LGSLAVSSLVLYARVGSLLDRLQIPIGTRGGARTQAEPIRALQFRKGN 75  
 7 MAALLGAVL-LVAQ---POLYPSRP-----ALGQQELLRKAGTLQD---- 45  
 Qy 17 ASKEQVRHLHDYQOLPKAIILIGYRKGETRALLEMLNLHPAVTKASQEIHFDDNDENYKGK 135  
 46 DVRDGAEPNGSACQLPQTIIIGYRKGETRALLEMLSLHPDVAAENHFFDWEHYSG 105  
 SQ 136 IEWYRKGMFSPSQQITKSPAYFTEEVPERIYKQNSKICLILIVTPPTRAIDYQ 195  
 Db 106 LGNYLSQMFPSWHQLVATEPKPAYTSKPVYVSNPSTRILLLRDPSRVLSDYQ 165  
 Qy 196 VLEGKERKNTKYKFEKLAIDPNTCEVNTKYKAVRTSLYTKHLERWTFPLQEFHVVG 255  
 Db 166 VYNHMQKHPKPSIEFLVRDG--RLNVDYKALNRSLYHVMQNWRFPFLRHIVDG 223  
 Qy 256 DRLIETPLPELOVKEFLNLPPIRSOYNLYENATRGFCYLRENNIFKCLAGSKGRTHPE 315



Page 9

XX	Penn SG, Hanzel DK, Chen W, Rank DR;	Qy	256 DRLITEPLPELQLVEKPINLPPISSQYNNATRQFYLRFNIIIFMKCLAGSKGRHPE 315
XX	WPI; 2002-114183/15.	Db	224 DRLIRPFPELQKVERFLKLSPQINNSNEYNKTRGFYCLR-DSGIDRCLHESKGRAHPQ 282
XX	Spatially-addressable set of single exon nucleic acid probes, used to measure gene expression in human lung samples -	Qy	316 VDPSTVTKLKKEFHFPNQKFYQITG 340
PT		Db	283 VDPKLINKLHEYFBPKKFELVG 307
XX			
PS	Claim 27; SEQ ID No 28566; 634pp; English.		
XX	The invention relates to a spatially addressable set of single exon nucleic acid probes for measuring gene expression in a sample derived from human lung comprising single exon nucleic acid probes having one of 12614 nucleic acid sequences mentioned in the specification, or their complements or the 12387 open reading frames derived from the 12614 probes. Also included are a microarray comprising the novel set of nucleic acid expressed in the human lung, measuring gene expression to a sample derived from human lung, comprising (a) contacting the array with a collection of detectably labeled nucleic acids derived from human lung mRNA, and (b) measuring the label detectably bound to each probe of the array; identifying exons in a eukaryotic genome, comprising (a) algorithmically predicting at least one exon from genomic sequences of the eukaryote; and (b) detecting specific hybridisation of detectably labeled nucleic acids from eukaryote lung mRNA, to a single exon probe, having a fragment identical to the predicted exon, the probe is included in the above mentioned microarray; assigning exons to a single gene, comprising (a) identifying exons from genomic sequence by the method above and (b) measuring the expression of each of the exons in several tissues and/or cell types using hybridisation to a single exon microarray having a probe with the exon, where a common pattern of expression of the exons in the tissues and/or cell types indicates that the exons should be assigned to a single gene, a peptide comprising one of 12011 sequences, mentioned in the specification, or encoded by the probes/open reading frames (ORE). The probes are used for gene expression analysis, and for identifying exons in a gene, particularly using human lung derived mRNA and for the study of lung diseases such as asthma, lung cancer, chronic obstructive pulmonary disease (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease, Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary haemangiopericytoma, lymphangioleiomyomatosis, pulmonary alveolar proteinosis, Kartagener syndrome, fibrocytic pulmonary dysplasia, primary ciliary dyskinesia, pulmonary hypertension and hyaline membrane disease. The present sequence is a peptide/protein encoded by a single exon probe of the invention.	RESULT 15 AYA17062 ID AYA17062 standard; Protein; 311 AA. XX AC AYA17062; DT 20-JUL-1999 (first entry) DS Mouse 3-OST-1 Protein. XX Heparan sulfate D-glucosaminyl 3-O-sulfotransferase; anticoagulant; KW saccharide; glycosaminoglycan; proteoglycan; antithrombin; 3-OST; KW heparan sulfate; thrombotic disorder; deep vein thrombosis; KW pulmonary embolism; coagulation enzyme inactivation. XX Mus musculus. OS XX PN WO992005-A2. XX PD 06-MAY-1999. XX PF 23-OCT-1998; 98WO-US22597. XX PR 31-OCT-1997; 97US-0065437. XX PR 24-OCT-1997; 97US-0062162. XX (MASSACHUSETTS INST TECHNOLOGY. PA PI Pritze LMS, Liu J, Rosenberg RD, Schwartz JJ, Shworak NW; PI Zhang L; XX WPI; 1999-312968/26. DR N-PSDB; AAX37245. XX Heparan sulfate D-glucosaminyl 3-O-sulfotransferase and related PT polynucleotides XX Claim 17; Page 71-72; 95pp; English.	
PS	Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at <a href="http://ftp.wipo.int/pub/published_pct_sequences">ftp.wipo.int/pub/published_pct_sequences</a> .	CC	The invention relates to nucleic acid molecules (AAX37245-X37250) encoding heparan sulfate D-glucosaminyl 3-O-sulfotransferases (3-OSTs). CC The 3-OST proteins can be used for 3-O-sulfating saccaride residues within a preparation of heparan sulfates (HS). CC The 3-OST proteins can be used for enriching the complex polysaccharides. 3-OST-1 can be used in a preparation of heparan sulfates (HSact). The 3-OSTs (optionally lacking enzymatic function) can be used to determine partial sequence information for CC antithrombin-binding fraction in a preparation of heparan sulfates (HS). CC The 3-OST proteins, genes and antibodies are also useful for diagnosis of disorders involving HS biosynthesis. 3-OSTs CC are especially used to generate anti-coagulant pentasaccharides, which may be used to treat thrombotic disorders such as deep vein thrombosis and CC pulmonary embolism. Coagulation enzyme inactivation by antithrombin is enhanced by complexing of antithrombin with endothelial cell surface CC proteoglycans. This is responsible for the non-thrombotic properties CC of blood vessels. The present sequence represents a mouse 3-OST-1. XX SQ Sequence 307 AA;
XX	Sequence 307 AA;	SQ	Sequence 311 AA;
Qy	Query Match 41.8%; Score 758; DB 23; Length 307;	Qy	Query Match 41.8%; Score 757.5; DB 20; Length 311;
Best Local Similarity 46.2%; Pred. No. 7e-67; Mismatches 88; Indels 24; Gaps 6;	Best Local Similarity 48.8%; Pred. No. 8e-67; Mismatches 54; Indels 19; Gaps 3;		
Matches 150; Conservative 63; Mismatches 88; Indels 24; Gaps 6;	Matches 143; Conservative 54; Mismatches 77; Indels 19; Gaps 3;		
Db	16 LGLSLAVGLLYLVARSGSLDRLQICPTEGRIGGARTQAEPPRLALGFRKGN 75	Db	16 VLEGKERRQTKYKPERLAIDPNTCEANTKYAVRTSTYRELLEWLKYFPLQEFFFVVDG 255
Qy	7 MAALLIGAVL-LVAQ----POLVPSRP-----AELSQQLLRKAGTLQD---- 45	Qy	76 ASKEQVRHLHDLYCQLPKRAIILIGVRGGTTRALIEMLNHPPAVYKASQEIHFENDENYKG 135
Db	46 DVRDGVADNGSAQQLPQTIIIGVRGGTTRALIEMLSUPDVAAEENVFHFWEEETSHG 105	Db	106 LGATLSONPFSWPHQLTYKTPAYFTSKVPERVYSMNPSTRILLIRDPSRVLSPTQ 165
Qy	136 IENYRKOMPFSYSPQQITIEKSPAYFTIEEPERIYKNSSIKULLIVREPTPTRAISDYTQ 195	Qy	196 VFTNHMQKHPKPSIEEFVNRDG--RLNDYKAQNRSVYHEVMQNLWFLPRLHHTVVDG 223
Db		Db	

```

Db      21 HPAAGPGIICQELLRKVILLPEDTGEIGIASNGSTQLPITIGVRKGETRALLEMLSL 80
QY      113 HPAVVKASOELHFFDIDENYKGGIEWYRKMPFSTPQQTTEKSAYFTBEVPERIYM 172
Db      81 HPDVAAAEENETHFFDMEBHSQSGLGMWLTQMPFESSPHOLTEKTPAYFTSPKVPERIHM 140
Db      173 NSSTKLILTYREPTTRAISDTQVLEGKERKNTKTYKEFEKLAIDNTCENTKYKAVENT 232
Db      141 NPTIRULLLRDPSEEVLSDTQVLYNHQHKPPIDLLMRG--RNLDYKALNES 198
Db      233 IYTAKHLERMKYFPIEQHYDGDRLITEPELQVEKPNLPRISQONLYNATRGF 292
Db      199 LYHAIRMLNWLRFFPLGHHTYDGDRLIRDSEPEIQLKVERTKLSPQINASNFTMKTKGF 258
QY      293 YCLRNNTIFNKLASSKGRTHPEVDPSVITLKRKFHPFQKFYQITGTLAW 345
Db      259 YCLR-DSGKORCLHESKGRAHPQVDBKLIDLXHEYFHEPKKPFKLUGRTFDW 310

```

Search completed: December 4, 2003, 16:44:55  
 Job time : 46 secs

GenCore version 5.1.6  
 Copyright (c) 1993 - 2003 Compugen Ltd.

## OM protein - protein search, using sw model

Run on: December 4, 2003, 16:43:02 ; Search time 20 Seconds  
 (without alignments)  
 1663.718 Million cell updates/sec

Title: PCT-US03-21094-2  
 Perfect score: 1814  
 Sequence: 1 MLEKQDAWLROKLILYUGSLA,.....FFHPPNOKFYQITGRTLNWP 346

Scoring table: BLOSUM62  
 Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96166682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR\_76;\*

1: Piri;\*

2: Piri;\*

3: piri;\*

4: Piri;\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	479	26.4	276	2	T33493 hypothetical prote
2	328	18.1	819	2	T29486 hypothetical prote
3	326.5	18.0	883	2	A92733 heparan sulfate - N-heparan sulfate
4	316.5	17.4	882	2	A42855 heparan sulfate - heparan sulfate
5	312.5	17.2	882	2	A57169 probable sulfonate - probable sulfonate
6	172	9.5	308	2	C59335 hypothetical prote
7	143.5	7.9	137	2	A94072 hypothetical prote
8	102	5.6	329	2	B96621 hypothetical prote
9	101.5	5.6	378	2	A90280 hypothetical prote
10	101.5	5.6	383	2	T51466 hypothetical prote
11	99	5.5	803	2	B36277 DNA-directed DNA p
12	99	5.5	853	2	S324682 probable transport
13	98.5	5.4	315	2	H70350 ADP-ribose synth
14	98.5	5.4	766	2	A86198 hypothetical prote
15	98.5	5.4	841	2	S324624 glycosidase
16	98.5	5.4	1495	2	S61023 hypothetical prote
17	98	5.4	470	2	A90083 hypothetical prote
18	98	5.4	502	2	S61928 probable membrane
19	97.5	5.4	692	2	S55849 nodulation protein
20	97	5.3	247	1	D24706 phosphoadenylyl-su
21	97	5.3	247	2	JC4531 alcohol sulfotrans
22	97	5.3	287	2	G93319 peptidylprolyl iso
23	97	5.3	430	2	B84946 rna binding protei
24	96	5.3	1279	2	T41389 DNA-directed RNA p
25	94	5.2	1361	2	T29486 probable membrane
26	94	5.2	796	2	S66769 trigger factor
27	93.5	5.2	442	2	CB4985 transcription init
28	93.5	5.2	631	2	G70188 hypothetical prote
29	93.5	5.2	1181	2	B64516

RESULT 1  
 T33493 hypothetical protein F40H3.5 - Caenorhabditis elegans

C;Species: Caenorhabditis elegans  
 C;Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999

C;Accession: T33493  
 R;Gattung, S.  
 submitted to the EMBL Data Library, October 1998

A;Description: The sequence of C. elegans cosmid F40H3 .

A;Reference number: 221358

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Cross-references: EMBL:AF098987; PIDN: AAC67432.1; GSPDB: GN000020; CBSP: F40H3.5

C;Genetics:

A;Gene: CBSP-F40H3.5

A;Map position: 2

A;Introns: 49/2; 109/3; 154/3; 227/3; 271/1

Query Match 26.4%; Score 479; DB 2; Length 276;

Best Local Similarity 40.9%; Pred. No. 6.3e-30; Mismatches 85; Indels 24; Gaps 7;

Matches 106; Conservative 44; Mismatches 85; Indels 24; Gaps 7;

QY 59 RALQFRKLHLHEFXKGNASKEQVRLHDLVQOLPKAIIGVRKGQTALLEMLNLHPAVK 118

Db 28 RNPFKKKWII-DLGKGEKS -----HLEKKFPSALIVGVRKGQTALLDATAHLHKVKRI 78

QY 119 ASQETHEFFDNDENYGKIGEWYRKMP-FSYXPOOITIEKSPAYFITEEVPERIYKNNSSIK 177

Db 79 VRRETHFFDS--NTLGFDWYRDRDNEPEVDNEIVTEKPAYFTNEHVPKRYENNPDMM 136

QY 178 LLLIVREPTRAISDVTQLEGKRKNTY-----YKFERKLADPTNCVNTKYKAVR 230

Db 137 LLLIVRHPYRVTFSDFQTYNNKLEQONKLPVLSEAFKTCNEAGE---KINNEYKEMT 192

QY 231 TSIVTKHLERWKYTFPEOPHYWDGRILTEPLBLQLVERKFLNLPPRISQNYLNFAATR 290

Db 193 NSLIDWHISKWLYKFDLKNFLFNGDVRANPLRELRKVEEFLGERSITPSQLVFDYNK 252

QY 291 GFYCLRNTEFLNKCLAGSK 309

Db 253 GFFCFR-KTITKVRCLGLSK 270

RESULT 2  
 T29486 hypothetical protein F00B4.6 - Caenorhabditis elegans

C;Species: Caenorhabditis elegans

C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jun-2000

C;Accession: T29486

## RESULT 1

T33493

hypothetical protein F40H3.5 - Caenorhabditis elegans

C;Sequence: Caenorhabditis elegans

C;Accession: T33493

R;Gattung, S.

submitted to the EMBL Data Library, October 1998

A;Description: The sequence of C. elegans cosmid F40H3 .

A;Reference number: T33493

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Cross-references: EMBL:AF098987; PIDN: AAC67432.1; GSPDB: GN000020; CBSP: F40H3.5

C;Genetics:

A;Gene: CBSP-F40H3.5

A;Map position: 2

A;Introns: 49/2; 109/3; 154/3; 227/3; 271/1

Query Match 26.4%; Score 479; DB 2; Length 276;

Best Local Similarity 40.9%; Pred. No. 6.3e-30; Mismatches 85; Indels 24; Gaps 7;

Matches 106; Conservative 44; Mismatches 85; Indels 24; Gaps 7;

QY 59 RALQFRKLHLHEFXKGNASKEQVRLHDLVQOLPKAIIGVRKGQTALLEMLNLHPAVK 118

Db 28 RNPFKKKWII-DLGKGEKS -----HLEKKFPSALIVGVRKGQTALLDATAHLHKVKRI 78

QY 119 ASQETHEFFDNDENYGKIGEWYRKMP-FSYXPOOITIEKSPAYFITEEVPERIYKNNSSIK 177

Db 79 VRRETHFFDS--NTLGFDWYRDRDNEPEVDNEIVTEKPAYFTNEHVPKRYENNPDMM 136

QY 178 LLLIVREPTRAISDVTQLEGKRKNTY-----YKFERKLADPTNCVNTKYKAVR 230

Db 137 LLLIVRHPYRVTFSDFQTYNNKLEQONKLPVLSEAFKTCNEAGE---KINNEYKEMT 192

QY 231 TSIVTKHLERWKYTFPEOPHYWDGRILTEPLBLQLVERKFLNLPPRISQNYLNFAATR 290

Db 193 NSLIDWHISKWLYKFDLKNFLFNGDVRANPLRELRKVEEFLGERSITPSQLVFDYNK 252

QY 291 GFYCLRNTEFLNKCLAGSK 309

Db 253 GFFCFR-KTITKVRCLGLSK 270

R;Stellyes, L.; Gattung, S.  
submitted to the EMBL Data Library, March 1996.

A;Description: The sequence of *C. elegans* cosmid F08B4.

A;Reference number: 220625

A;Accession: T29486

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-819 <STB>

A;Cross-references: EMBL:US2005; PIDN:AA837730\_1; GSPPDB:GN00022; CEESP:F08B4\_6

C;Genetics:

A;Gene: CEESP:F08B4\_6

A;Experimental source: strain Bristol N2; clone F08B4

C;Superfamily: Caenorhabditis elegans hypothetical protein F08B4\_6

A;Map position: 4

A;Introns: 47/2; 142/3; 259/3; 360/3; 465/3; 526/3; 561/1; 656/3; 726/3

Query Match Score 328; DB 2; Length 819;

Best Local Similarity 31.2%; Pred. No. 1..6e-17;

Matches 80; Conservative 48; Mismatches 118; Indels 10; Gaps 4;

Qy 88 CQLPKAIIIGVRKGITGTRALLEMINLHP-----AVRKPAHEDNGRGENYR 141

Db 547 KSLPDLIIGPQXTGSTALASPLHNTSQTNPVPSSEEVQFP-GGQNYLKGVEMMS 605

## RESULT 4

Qy 142 KMPFSPYQQITIEKSPAYITEEVPERLYKNSIKLILIVREPTTAISDYZQLEGKE 201

Db 606 NFSS--STVTFERSATYFDNNSAPKQASLYPHAKIVTILLONPAQAYSWFQHILAHED 663

Qy 202 RQNKTYKPFKLAIDPNTCEVNTKYKAVRTSYTKHLERMLKYPFIEQFHVDGDLITE 261

Db 664 PVAITAGSLEVLDSNSTSKVKRORCTSGGRVHLTLEHFSLSQMMIFDSDELMKM 723

Qy 262 FPEFLQVEKFNLNLPTRISQNYVNAFRGTCFLRENTIINCLASSKGRIHPEVDFSVI 321

Db 724 PPTVNLNSKWLDPPEPETVIRYSFSKGFFHC-RLLDGTKTKOLGESKGKXPEMPENLR 782

Qy 322 TKLRKFFPHPNOKFYQ 337

Db 783 RKLDKIFKFLSDNSALYK 798

## RESULT 5

Qy 49733 Heparan sulfate-glycosamine N-sulfotransferase (EC 2.8.2.8) - mouse

N;Alternative names: N-heparan sulfate sulfotransferase (N-HSST)

N;Contains: desulfoheparin sulfate transferase (EC 2.8.2.8); Glycosaminoglycan N-acetylgluc

C;Date: 03-May-1994 #sequence\_revision 03-May-1994 #text\_change 03-Jun-2002

C;Accession: A49733; A53736

R;Orellana, A.; Hirschberg, C.B.; Wei, Z.; Swiedler, S.J.; Ishihara, M.

J. Biol. Chem., 269, 2270-2276, 1994

A;Title: Molecular cloning and expression of a glycosaminoglycan N-acetylglucosaminyl N-

A;Reference number: A49733; PMID:8293485

A;Accession: A49733

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1..883 <OPE>

A;Cross-references: GB:U02304; PIDN:AA52137.1; PID:9457944

R;Eriksson, I.; Sandbeck, D.; Ek, B.; Lindahl, U.; Kjelleen, L.

J. Biol. Chem., 269, 10438-10443, 1994

A;Title: cDNA cloning and sequencing of mouse mastocytoma glucosaminyl N-deacetylaselase/N-

A;Reference number: A53736; PMID:94193735; PMID:8144627

A;Accession: A53736

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1..117 'V'..119-395..397-883 <ERI>

A;Cross-references: GB:X7585; PIDN:9474430; PID:CA53479.1; PID:9474431

C;Superfamily: Caenorhabditis elegans hypothetical protein F08B4\_6

C;Keywords: carboxylic ester hydrolase; glycoprotein; sulfotransferase; transmembrane

Query Match Score 18.0%; DB 2; Length 883;

Best Local Similarity 31.2%; Pred. No. 2..3e-17;

Matches 86; Conservative 48; Mismatches 125; Indels 17; Gaps 7;

Qy 77 SKEQVRLHDLYVQQLPKAIIIGVRKGCGTRALLEMINLHPAVVKA----SQBTIHFDNDNE 130

Db 595 SKEKT----CDRJPFLIGSPQKTTTAHFELSHPATSSPSSTPEIOPF-NGP 648

Qy 131 NYKGKIEWYRK---KMPFSPYQQITIEKSPAYITEEVPERLYKNSIKLILIVREPT 188

Db 649 NYHKGIDWMDDEPPVSNASTDFLFFERSATYDFSEVTPRGAALLERAKITLVLNPADR 708

Qy 189 AISDTY-QVLEGKE-RKQKTYKPFKLAIDPNTCEVNTKYKAVRTSYTKHLERMLKYPF 246

Db 709 AYSWYHQRAHGDPIALNNTFYQVASSQAPLILRSQNRCYPSYSTHLQRWTTYYP 768

Qy 247 IBOQHVVTDGDELITELPLOLVEFLNLNLPPLRISLQNLQFLGIPFQLEGKTRCL 828

Db 769 SGQLIIMDGQELRVNPAASMEILIQFLGIPFQLEGKTRCL 305

Qy 306 AGSKGRHPEVDPSTIKRKFPHPNOKFYQITGR 341

Db 829 GRSKGRHPEVDPSTIKRKFPHPNOKFYQITGR 864

## RESULT 5

Qy 77 SREQVRLHDLYVQQLPKAIIIGVRKGCGTRALLEMINLHPAV----VKA5OBHFDNDNE 130

C;Accession: A42855

R;Hashimoto, Y.; Orellana, A.; Gil, G.; Hirschberg, C. B.

J. Biol. Chem., 267, 15744-15750, 1992

A;Title: Molecular cloning and expression of rat liver N-heparan sulfate sulfotransf

A;Reference number: A42855; PMID:92348437; PMID:1379236

A;Accession: A42855

A;Status: preliminary

A;Molecule type: mRNA; protein

A;Experimental source: liver

A;Note: sequence inconsistent with the nucleotide translation

A;Note: sequence extracted from NCBI block protein (NCIN:109827; NCIBP:109828)

C;Superfamily: Caenorhabditis elegans hypothetical protein F08B4\_6

Query Match Score 17.4%; DB 1.4e-16; Best Local Similarity 32.2%; Pred. No. 1..4e-16; Matches 93; Conservative 38; Mismatches 127; Indels 31; Gaps 10;

Qy 77 SREQVRLHDLYVQQLPKAIIIGVRKGCGTRALLEMINLHPAV----VKA5OBHFDNDNE 130

C;Accession: A42855

R;Hashimoto, Y.; Orellana, A.; Gil, G.; Hirschberg, C. B.

J. Biol. Chem., 267, 15744-15750, 1992

A;Title: Molecular cloning and expression of rat liver N-heparan sulfate sulfotransf

A;Reference number: A42855; PMID:92348437; PMID:1379236

A;Accession: A42855

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1..882 <HAS>

A;Note: sequence inconsistent with the nucleotide translation

A;Note: sequence extracted from NCBI block protein (NCIN:109827; NCIBP:109828)

C;Superfamily: Caenorhabditis elegans hypothetical protein F08B4\_6

Qy 77 IBOQHVVTDGDELITELPLOLVEFLNLNLPPLRISLQNLQFLGIPFQLEGKTRCL 828

C;Accession: A42855

R;Hashimoto, Y.; Orellana, A.; Gil, G.; Hirschberg, C. B.

J. Biol. Chem., 267, 15744-15750, 1992

A;Title: Molecular cloning and expression of rat liver N-heparan sulfate sulfotransf

A;Reference number: A42855; PMID:92348437; PMID:1379236

A;Accession: A42855

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1..883 <OPE>

A;Cross-references: GB:U02304; PIDN:AA52137.1; PID:9457944

R;Eriksson, I.; Sandbeck, D.; Ek, B.; Lindahl, U.; Kjelleen, L.

J. Biol. Chem., 269, 10438-10443, 1994

A;Title: cDNA cloning and sequencing of mouse mastocytoma glucosaminyl N-deacetylaselase/N-

A;Reference number: A49733; PMID:94193735; PMID:8144627

A;Accession: A49733

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1..117 'V'..119-395..397-883 <ERI>

A;Cross-references: GB:X7585; PIDN:9474430; PID:CA53479.1; PID:9474431

C;Superfamily: Caenorhabditis elegans hypothetical protein F08B4\_6

C;Keywords: carboxylic ester hydrolase; glycoprotein; sulfotransferase; transmembrane

Query Match Score 18.0%; DB 2; Length 883;

Best Local Similarity 31.2%; Pred. No. 2..3e-17;

Matches 86; Conservative 48; Mismatches 125; Indels 17; Gaps 7;

C;Species: Homo sapiens (man)

C;Date: 23-Feb-1996 #sequence\_revision 23-Feb-1996 #text\_change 31-Mar-2003

## RESULT 5

A57169

The heparan sulfate]-glucosamine N-sulfotransferase (EC 2.8.2.8) - human

N;Alternate names: N-heparan sulfate sulfotransfase (N-HSST)

N;Contains: desulfoleparin sulfate sulfotransfase (EC 2.8.2.8); glycosaminoglycan N-acetylgluc

C;Species: Homo sapiens (man)

C;Date: 23-Feb-1996 #sequence\_revision 23-Feb-1996 #text\_change 31-Mar-2003

C;Accession: A57169; PID:G02129; OMIM:601581  
 R;Dixon, J.; Loftus, S.K.; Gladwin, A.J.; Scambler, P.J.; Wasmuth, J.J.; Dixon, M.J.  
 Genomics 26, 239-244, 1995  
 A;Title: Cloning of the human heparan sulfate-N-deacetylase/N-sulfotransferase gene from  
 A;Reference number: A57169; MUID:55324914; PMID:76014418  
 A;Accession: A57169  
 A;Molecule type: mRNA  
 A;Residues: 1-882 <DIX>  
 A;Cross-references: GB:U8910; PIDN:MAA5281.1; PID:9976372  
 A;Note: authors translated the codon AAG for residue 42 as Leu, AAA for residue 110 as Ile  
 R;Humphries, D.E.  
 submitted to the EMBL Data Library, September 1995  
 A;Reference number: G09202  
 A;Accession: G09229  
 A;Status: preliminary; translated from GB/EMBL/DDJB  
 A;Molecule type: mRNA  
 A;Residues: 1-882 <HDM>  
 A;Cross-references: EMBL:U36600; PID:gi1036796; PIDN:ACC27354.1; PID:gi1036797  
 R;J.Bell, T.L.  
 submitted to the EMBL Data Library, December 1994  
 A;Reference number: G07229  
 A;Accession: G01581  
 A;Status: preliminary; translated from GB/EMBL/DDJB  
 A;Molecule type: mRNA  
 A;Cross-references: EMBL:U17970; PIDN:gi641163; PID:9841164  
 C;Genetics:  
 A;Gene: GDB:HSSST; NT\_031  
 A;Cross-references: GDB:593916; OMIM:600853  
 A;Map position: 5q3-5q3.1  
 C;Function: <DAc>  
 A;Description: as glycosaminoglycan N-acetylglucosaminyl N-deacetylase, hydrolyzes the N-  
 A;Pathway: heparan sulfate biosynthesis  
 C;Function: <GST>  
 A;Description: as desulfoglycosaminoglycan sulfotransferase, catalyzes the formation of glucosamin  
 enoine-5-phosphosulfate and converting it to adenosine 3',5'-bisphosphate  
 C;Superfamily: Caenorhabditis elegans hypothetical protein F08B4.6  
 C;Keywords: Glycoprotein; Golgi apparatus; hydrolase; sulfotransferase; transmembrane protein  
 F117/Domain: intracellular #status predicted <INT>  
 P18-39/Domain: transmembrane #status predicted <TRM>  
 F40-882/Domain: trans-Golgi network luminal #status predicted <LUM>  
 F231,351,401,667/Binding site: carbohydrate (Asn) (covalent) #status predicted  
 Query Match Score 312.5; DB 2; Length 822;  
 Best Local Similarity 30.6%; Pred. No. 2.8e-16;  
 Matches 92; Conservative 41; Mismatches 113; Indels 55; Gaps 11;  
 Db 596 SKEKT----CDRFPKLLIIPQKIGTATYLFLGMHPDSSNYSSETEEIQFF-NGH 649  
 Qy 77 SKEQYRLHDLYQQLPKAIIGIVRKGGTRALLEMLNIHPPAV-----VKASOEHFFFDNDE 130  
 Db 650 NYHKGDWYMFPEFPI--PSNTSDFYFEKSANYFDSEVAPRAAALLPKAVLTILINPA 707  
 Query Match Score 312.5; DB 2; Length 822;  
 Best Local Similarity 30.6%; Pred. No. 2.8e-16;  
 Matches 92; Conservative 41; Mismatches 113; Indels 55; Gaps 11;  
 Db 596 SKEKT----CDRFPKLLIIPQKIGTATYLFLGMHPDSSNYSSETEEIQFF-NGH 649  
 Qy 131 NYGKGIGEWYRKCKMPFSYPOQT---IEKSPAYFITEVPERIYKNNSSKILLIVREPT 186  
 Db 650 NYHKGDWYMFPEFPI--PSNTSDFYFEKSANYFDSEVAPRAAALLPKAVLTILINPA 707  
 Query Match Score 312.5; DB 2; Length 822;  
 Best Local Similarity 30.6%; Pred. No. 2.8e-16;  
 Matches 92; Conservative 41; Mismatches 113; Indels 55; Gaps 11;  
 Db 596 SKEKT----CDRFPKLLIIPQKIGTATYLFLGMHPDSSNYSSETEEIQFF-NGH 649  
 Qy 187 TRAISDYTQVLEGKERKNTKYKFEKLAIDNT-----CEVNNTKKYKAVRTSI- 233  
 Db 708 DRAYS-----WYQHORAHDDPVALKYTFEHVITYTAGSDASSKLRLAQNRCL 752  
 Qy 234 ----YTKHLEWKYKPIEOPHIVDPBDRLTEPLPLPQLVYKEFLNLPPRISQY-NLYFNA 288  
 Db 753 VPGWYATHIERWNLSAATHANQIILVLDGKLLREPAXYMDMVKQLFGYNTIDYHTKTLAFDP 812  
 Qy 289 TRGFYCILRFNTLIFENKCLAGSKGRHPEVDPSVITKURKFHFPN--OKTYQITGRTL-N 344  
 Db 813 KRGFWWCOLLEGKTKCIGKSKGRKYPEMDLDSRAFLKDYRDHNTELSKLYKNGQTLPT 872  
 Query Match Score 312.5; DB 2; Length 822;  
 Best Local Similarity 30.6%; Pred. No. 2.8e-16;  
 Matches 92; Conservative 41; Mismatches 113; Indels 55; Gaps 11;  
 Db 596 SKEKT----CDRFPKLLIIPQKIGTATYLFLGMHPDSSNYSSETEEIQFF-NGH 649  
 Qy 345 W 345  
 Db 873 W 873

RESULT 6  
 C9535  
 probable sulfotransferase protein [imported] - Sinorhizobium meliloti (strain 1021) mag  
 C;Species: Sinorhizobium meliloti  
 C;Date: 24-Aug-2001 #text\_change 30-Sep-2001  
 C;Accession: C95935  
 R;Finan, T.M.; Weidner, S.; Wong, K.; Buhmester, J.; Chain, P.; Vorholter, F.J.; Hernai  
 Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001  
 A;Title: The complete sequence of the 1,683-kb pSymB megaplasmid from the N2-fixing end  
 A;Reference number: A95842; MUID:21396508; PMID:11481431  
 A;Accession: C95935  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-308 <KUR>  
 A;Cross-references: GB:AL591985; PIDN:CAAC49147.1; PID:gi15140632; GSPDB:GN00167  
 A;Experimental source: strain 1021, megaplasmid pSymB  
 R;Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Ampe, F.; Abola, P.; Vorholter, F.; Barloy-Hubler  
 pela, D.; Chain, P.; Cowie, A.; Fedderspiel, N.A.; Fisher, R.F.  
 L.; Hyman, R.W.; Jones, T.  
 Science 293, 668-672, 2001  
 A;Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lealaure  
 hebault, P.; Vandembol, M.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.  
 A;Title: The composite genome of the legume symbiont Sinorhizobium meliloti.  
 A;Reference number: A96039; MUID:21368234; PMID:11474104  
 A;Contents: annotation  
 C;Genetics:  
 A;Gene: Smb21243  
 A;Genome: Plasmid  
 Query Match Score 9.5%; DB 2;  
 Best Local Similarity 24.5%; Pred. No. 6.3e-06;  
 Matches 65; Conservative 38; Mismatches 120; Indels 42; Gaps 7;  
 Db 85 DLVOOPLKA-----IIIGIVRKGGTRALLEMLNIHPPAVKASOEHFFNDENYKGIE 137  
 Db 22 EAIEPKRNEMRDIDFLIGATKSATWVQSLQDGPGLMPDPBLHYF-SRYERGDE 79  
 Qy 138 WYRKMKPFSYPOQITIEKSPAYFITEVEPRTYKNSKSKILLIVREPTTRALSDYTVL 197  
 Db 80 WYLEHFAQGEHRRURGEKNSNTMDVPEAERIKELPPEARLIAHVRNPVDRASDXCMLY 139  
 Qy 198 EGKE--RKRKTYKKEFLAKAIDPNTCBAVNTKYKAVRTSIVTKHLRKLWKLKYPFIEOPHVVDG 255  
 Db 140 RRAFYGRDIAQ----LDPRO--GAGGRFENGHYYQQLCYLDPFPAEQIVLWV 189  
 Qy 256 DRLETPLPLOLVKEFLNLPPRSQNLNFATRFGYCFIRNTIFNKCLAGSKGRIRHPE 315  
 Db 190 EDLKLDRAGLARV-----RGFLGLEADVPLPLAKKVKDKESEPV 229  
 Query Match Score 9.5%; DB 2;  
 Best Local Similarity 24.5%; Pred. No. 6.3e-06;  
 Matches 65; Conservative 38; Mismatches 120; Indels 42; Gaps 7;  
 Db 316 VDPSTYTKRKEPKFEPHQKQFYQITG 340  
 Db 230 VNPT-LRRLRPLKPVAAFRONTG 253  
 RESULT 7  
 A84072 hypothetical protein BH3377 [imported] - Bacillus halodurans (strain C-125)  
 C;Species: Bacillus halodurans  
 C;Date: 01-Dec-2000 #text\_change 15-Jun-2001  
 C;Accession: A84072  
 R;Takami, H.; Nakatone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fuji, F.; Hir  
 Nucleic Acids Res. 28, 4317-4331, 2000  
 A;Title: Complete genome sequence of the alkaliophilic bacterium Bacillus halodurans and  
 A;Reference number: A83650; MUID:20512582; PMID:11058132  
 A;Accession: A84072  
 A;Molecule type: DNA  
 A;Residues: 1-137 <STO>  
 A;Cross-references: GB:AP001518; GB:BA000004; NID:gi10175792; PIDN:BAB07096.1; GSPDB:GN  
 A;Experimental source: strain C-125  
 C;Genetics:  
 A;Gene: BH3377

Query Match 7.9%; Score 142.5; DB 2; Length 137;  
 Best Local Similarity 25.0%; Pred. No. 0.00036; DB 28; Mismatches 59; Indels 27; Gaps 5;  
 Matches 38; Conservative 28; Mismatches 59; Indels 27; Gaps 5;  
 C:Date: 24-May-2001 #sequence\_revision 24-May-2001 #text\_change 24-May-2001  
 C:Accession: A90280  
 R:She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Avayez, M.J.; Chan  
 Jong, I.; Jeffries, A.C.; Kozares, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder,  
 arrett, R.A.; Rogan, M.A.; Sensen, C.W.; Van der Oost, J.;  
 submitted to Genbank April 2001  
 A:Description: Sulfolobus solfataricus complete genome.  
 A:Reference number: A99139  
 A:Accession: A90280  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-378 <CUR>  
 A:Cross-references: GB:AE006641; PIDN:AAK41496.1; GSPDB:GN00155  
 C:Genetics:  
 A:Gene: SSO1258

Query Match 5.6%; Score 101.5; DB 2; Length 378;  
 Best Local Similarity 22.5%; Pred. No. 2.6%; DB 21; Mismatches 63; Indels 79; Gaps 21;  
 Matches 81; Conservative 63; Mismatches 137; Indels 79; Gaps 21;  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Accession: H96621  
 C:Status: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
 R:Theologis, A.; Becker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; anzen, N.; Hughes, B.; Ruizar, L.;  
 Nature 408, 816-820, 2000  
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.; Li, J.H.; Li, Y.; Liu, S.X.; Liu, Z.A.; Luos, J.S.; Maiti, R.; Marziali, R.; Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.;  
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shim, P.; Southwick, A.M.; Sun, H.; Tallon, Kee, M.; Wu, D.; Yu, J.; Fraser, R.W.; Venner, J.C.; Davis, R.W.;  
 A:Title: Sequence and analysis of chromosome 1 of the Plant Arabidopsis.  
 A:Reference number: A86141; PMID:21016719; PMID:11130712  
 A:Accession: H96621  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Cross-references: GB:AB005173; PIDN:AAD39322.1; GSPDB:GN00141  
 C:Genetics:  
 A:Gene: F23H11.11  
 A:Map position: 105 QKINKETRRRLYYFKPFNDQLYALINKNGFW 136

RESULT 8

Query Match 4.0%; Score 99.5; DB 2; Length 378;  
 Best Local Similarity 22.5%; Pred. No. 2.6%; DB 21; Mismatches 63; Indels 79; Gaps 21;  
 Matches 81; Conservative 63; Mismatches 137; Indels 79; Gaps 21;  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Accession: H96621  
 C:Status: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
 R:Theologis, A.; Becker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; anzen, N.; Hughes, B.; Ruizar, L.;  
 Nature 408, 816-820, 2000  
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.; Li, J.H.; Li, Y.; Liu, S.X.; Liu, Z.A.; Luos, J.S.; Maiti, R.; Marziali, R.; Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.;  
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shim, P.; Southwick, A.M.; Sun, H.; Tallon, Kee, M.; Wu, D.; Yu, J.; Fraser, R.W.; Venner, J.C.; Davis, R.W.;  
 A:Title: Sequence and analysis of chromosome 1 of the Plant Arabidopsis.  
 A:Reference number: A86141; PMID:21016719; PMID:11130712  
 A:Accession: H96621  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Cross-references: GB:AB005173; PIDN:AAD39322.1; GSPDB:GN00141  
 C:Genetics:  
 A:Gene: F23H11.11  
 A:Map position: 105 QKINKETRRRLYYFKPFNDQLYALINKNGFW 136

RESULT 10

Query Match 5.6%; Score 102; DB 2; Length 329;  
 Best Local Similarity 21.8%; Pred. No. 2%; Mismatches 38; Indels 80; Gaps 13;  
 Matches 58; Conservative 38; Mismatches 90; Indels 80; Gaps 13;  
 C:Species: Arabidopsis thaliana (mouse-year cress)  
 C:Accession: T51466  
 C:Status: 18-Aug-2000 #sequence\_revision 18-Aug-2000 #text\_change 18-Aug-2000  
 R:Sato, S.; Nakamura, Y.; Kaneko, T.; Kato, T.; Asamizu, E.; Kotani, H.; Tabata, S.; Me submitted to the Protein Sequence Database, August 2000  
 A:Accession: T51466  
 A:Map position: 5  
 A:Molecule type: DNA  
 A:Cross-references: EMBL:Al391151  
 A:Experimental source: cultivar Columbia; BAC clone K10A8  
 C:Genetics:  
 A:Map position: 5  
 A:Introns: 45/2; 74/1; 94/3; 125/2; 167/2; 197/3; 233/3; 366/1  
 A:Note: K10A8\_100

Query Match 5.6%; Score 101.5; DB 2; Length 383;  
 Best Local Similarity 21.8%; Pred. No. 2.6%; DB 20; Mismatches 60; Indels 105; Gaps 20;  
 Matches 83; Conservative 60; Mismatches 133; Indels 105; Gaps 20;

Query Match 1 MLFKQAOAWLROKILVNL---GSLAVSLLYLVARGSLSLDRLOQICIEGRIGGARTOAEEF 56  
 Best Local Similarity 21.8%; Pred. No. 2.6%; DB 20; Mismatches 60; Indels 105; Gaps 20;  
 Matches 83; Conservative 60; Mismatches 133; Indels 105; Gaps 20;

Query Match 1 MAQKMBEIQKRLRLSYPRANAPAQLLF---AGMERYALLEWLFKKLG--DQSPP 53  
 Best Local Similarity 21.8%; Pred. No. 2.6%; DB 20; Mismatches 60; Indels 105; Gaps 20;

RESULT 9

A9280 hypothetical protein SSO1258 [imported] - Sulfolobus solfataricus  
 C:Species: Sulfolobus solfataricus

Qy 57 PLRALQPFGLLHFRIKGNASKEQVRLHDLVQ-----QLPKAIIIGVRKCGT-RAILEM 109  
 Db 54 SQNLQCDGQDVE-----ETVRQQLIATAKFLGITPTDIAIQGTYEDRME 105  
 Qy 110 LNLHPAVYKQASQEIHFDN-----DENYKGIEW-----RKKMPFSS-----YPQQITIE 154  
 C;Species: Saccharomyces cerevisiae  
 C;Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 21-Jul-2000  
 C;Accession: S34682; S38020; S38021; S38022; S38023; S38024  
 Db 106 LRNIVDIVEASL---FSBNQEWNSIDEQVADILQIAEROSLIFSEECKLFPADYQIQ 162  
 R;Wieman, S.; Voss, H.; Schwaeer, C.; Rupp, T.; Stegemann, J.; Grothue:  
 submitted to the EMBL Data Library, July 1993  
 A;Description: Sequencing and analysis of 51.5 kilobases on the left arm of chromosome :  
 A;Accession number: S34679  
 A;Accession: S34679  
 A;Molecule type: DNA  
 A;Residues: 1-853 <WIE>  
 A;Cross-references: EMBL:X74151; NID:g450365; PID:g395237  
 R;Wiemann, S.; Voss, H.; Schwaeer, C.; Rupp, T.; Grothue, D.; Sensen, C.; Stegemann, J.  
 A;Reference number: S37825  
 A;Submitted to the Protein Sequence Database, March 1994  
 A;Accession: S38020  
 A;Molecule type: DNA  
 A;Residues: 1-53 <W12>  
 A;Cross-references: EMBL:Z228188; NID:g486332; PID:9486333; MIPS:YKL188C  
 A;Experimental source: strain S288C  
 R;Maia e Silva, A.; Bossier, P.; Vilela, C.; Fernandes, L.; Soares, H.; Guerreiro, P.;  
 A;Reference number: S38024  
 A;Submitted to the Protein Sequence Database, March 1994  
 A;Accession: S38025  
 A;Molecule type: DNA  
 A;Residues: 1-53 <W12>  
 A;Cross-references: SGD:S0001671; MIPS:YKL188C  
 A;Gene: SGD:PX22  
 A;Cross-references: SGD:S0001671; MIPS:YKL188C  
 A;Map Position: 11L  
 C;Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homology  
 P:498-722/Domain: ATP\_P-Loop; peroxisome; purine nucleotide binding; transmembrane protein  
 F:505-512/Region: nucleotide-binding motif A (P-Loop)  
 F:511/Binding site: ATP/GTP (lys) #status predicted

RESULT 11  
 B56277  
 DNA-directed DNA polymerase (EC 2.7.7.7) II - Pyrodictium occultum  
 C;Species: Pyrodictium occultum  
 C;Date: 03-Oct-1995 #sequence\_revision 03-Oct-1995 #text\_change 20-Jun-2000  
 C;Accession: B56277  
 R;Uemori, T.; Ishino, Y.; Doi, H.; Kato, I.  
 J. Bacteriol. 177: 2164-2177, 1995  
 A;Title: The hyperthermophilic archaeon Pyrodictium occultum has two alpha-like DNA polymerases situated on the left arm of chromosome XI codes for a homologous gene YKL741  
 A;Reference number: S44318  
 A;Accession: SGD:S0001671  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-803 <WEM>  
 A;Cross-references: GB:D38574; NID:g9807829; PID:9807830  
 C;Keywords: herpesvirus DNA-directed DNA Polymerase  
 C;Superfamily: nucleotidyltransferase

Query Match Score 99; DB 2; Length 803;  
 Best Local Similarity 18.6%; Pred. No. 11; Mismatches 44; Indels 126; Gaps 14;  
 Matches 63; Conservative 44; Mismatches 106; Indels 125; Gaps 12;

Qy 81 VRLHDLYQQLPK-----AIIIGTRKGGRFALLMLNHPAVVKASQEIHFFNDEN 131  
 Db 295 VDLYDAEPMPEIKMKTLVEVABLYGMKSKSERVLEWRIP-----EYWD-DEK 343  
 A;Cross-references: GB:D38574; NID:g9807829; PID:9807830  
 C;Keywords: nucleotidyltransferase

Query Match Score 99; DB 2; Length 853;  
 Best Local Similarity 21.1%; Pred. No. 12; Mismatches 46; Indels 125; Gaps 12;  
 Matches 58; Conservative 46; Mismatches 125; Indels 125; Gaps 12;

Qy 132 YGKGIEW-----RKKMPFSSPYQQITIEPS-----AY 159  
 Db 344 KRQRLYALDDVRYATYGLAEKMLPPAI-OLSTVTVGPLDQVGAMGVGERLEWLMRAAY 402  
 Qy 160 FITTEVPERYKMNSSIKLLIVREPTPRTSDFYDQVLGKERKRNKTYKF----EKLAI 215  
 Db 403 DMNELLPNRVERRGESTYKGAVVTK- PLKGTYHENVVLDSSMYSPIIMXNVGPDFTIV 460  
 Qy 216 DENTC-----EVNTYKAVRTSILYKHLERMLKYPFLEQFHVDGDRLITEPLPE 265  
 Db 461 DSBCCPKYGCGYVAPEYGHFRRSPPGFFKTVLENLKL-----R 500  
 Db 403 DMNELLPNRVERRGESTYKGAVVTK- PLKGTYHENVVLDSSMYSPIIMXNVGPDFTIV 460  
 Qy 266 QLOVEKELNLNPRLPRISQONLY-----FRATGF-----YCLRF-----N 298  
 Db 501 RQVKEYKMKFEPDSPEIRLDERQKALKVLANASYGYMGWSHAWTYCKRCAEAATWGRN 560  
 Qy 299 IIFN-----KCHAGSKSRGIHPEVPDVPSVTKLRF 327  
 Db 561 LILTAIEYARKLGLKVIYGGDTSLSFVYYDKERKEVKLIEF 599  
 Qy 257 RLITDPLPELQLVERKFLNPDPRIQSQONLYFNATRG 291  
 Db 771 NIIRKSETNLNLFKEKIED--PKTSKSNALFNANK 803

2

D	315	KFGCRDDBVRINPDAFPDKVIAQNDTHPLAIPENIRLVDLERLDWDKAHDVTYRKTAY	374
Q	235	TKH---LERWKYFPLEQFHVVGDGLITEPLP-ELQV---EKFLN---LPPR 278	
Q	375	TNHTVILPEALERW---IV-----HMETULPRHOIYEVNOREFLNRVAAFPGD 421	
D	279	ISQVNLYFNATRGFYCLRFNIIFPKCLASKS---GRIHPEVDSSVITLKRKF---H 329	
Q	422	VDRLLRNSLVEES-AVKRINMA-HLCIAGSHAVNGVARIHSEELKKTIK-DFFBLEPH 477	
Q	330	PFNQKPYOITGR 341	
D	478	KFONKNTGIPTR 489	

Search completed: December 4, 2003, 16:46:05  
Job time : 28 secs

Result No.	Score	Query Match	Length	DB ID	Description
1	9.5	18.4	883	1 HSS2_HUMAN	P52849 homo sapien
2	326.5	18.0	883	1 HSS2_MOUSE	P52850 mus musculus
3	316.5	17.4	882	1 HSS1_RAT	Q02353 ratus norvegicus
4	312.0	17.2	882	1 HSS1_HUMAN	P52848 homo sapien
5	10.4	5.7	247	1 NOHA_PHIME	P06237 rhizobium m
6	9.9	5.5	853	1 PXA2_YEAST	P34230 saccharomyces cerevisiae
7	98.5	5.4	841	1 PHS2 RAT	D09812 ratus norvegicus
8	97.5	5.4	472	1 PAP1_FONEY	Q9J5B8 fowipox virus
9	97.5	5.4	692	1 YUH9_YEAST	P40358 saccharomyces cerevisiae
10	9.7	5.3	247	1 NODH_RHIMB	P06236 rhizobium m
11	9.7	5.3	286	1 SUBB_CAYPO	P52841 carica porcata
12	9.7	5.3	410	1 SORA_BUCAI	P57240 buchnera apicalis
13	9.7	5.3	1043	1 AG01_ARATH	Q04379 arabidopsis thaliana
14	9.5	5.2	803	1 GYRB_BUCAPI	P29435 buchnera apicalis
15	94.5	5.2	1361	1 RPOD_SPIGUL	P11704 spinacia officinalis
16	9.4	5.2	796	1 DECI_YEAST	Q12387 saccharomyces cerevisiae
17	9.3	5.2	442	1 TIG_BUCAPI	P67546 buchnera apicalis
18	93.5	5.2	631	1 RPSD_BORBU	P52323 borrelia burgdorferi
19	93.5	5.2	1181	1 YY02_METTA	O60301 methanococcus marcusii
20	91.5	5.1	662	1 UVRE_THEMNAER	Q9J8M thermoaerophilus
21	92.5	5.1	585	1 YH70_SYNTH3	P73627 synechocystis sp
22	92	5.1	1125	1 MFD_BORBU	O51568 borrelia burgdorferi
23	9.2	5.1	1603	1 STU1_YEAST	P06125 caenorhabditis elegans
24	9.2	5.1	2004	1 VIT5_CAEEL	Q84dJ2 yersinia pestis
25	91.5	5.0	761	1 YA88_AQUAB	0671178 aquifex aeolicus
26	91.5	5.0	841	1 PHS2_HUMAN	P11217 homo sapiens
27	91.5	5.0	1065	1 SMCS5_SCHPO	Q13710 schizosaccharomyces pombe
28	91.5	5.0	1513	1 STU1_YEAST	P38198 saccharomyces cerevisiae
29	91.5	5.0	1603	1 VIT5_CAEEL	Q84dJ2 caenorhabditis elegans
30	91.5	5.0	1603	1 VIT4_CAEEL	P18947 caenorhabditis elegans
31	91	5.0	881	1 YEBT_HAENIN	P44288 haemophilus suis
32	91	5.0	1501	1 CDR3_CANAL	O26290 candida albicans
33	91	5.0	1597	1 CTRD_MOUSE	P49025 mus musculus

ALIGNMENTS					
<b>RESULT 1</b>					
HSS2_HUMAN	ID: HSS2_HUMAN	STANDARD;	PRT;	883 AA.	
AC: P52849;	DT: 01-OCT-1996 (Rel. 34, Created)				
	DT: 01-OCT-1996 (Rel. 34, Last sequence update)				
	DT: 15-SEP-2003 (Rel. 42, Last annotation update)				
DB: HSST) (N-heparin sulfate N-sulfotransferase (EC 2.8.2.-))	DB: Heparin sulfate N-deacetylase/N-sulfotransferase (Glucosaminyl N-deacetylace/N-sulfotransferase)	(Glucosaminyl N-deacetylace/N-sulfotransferase).			
GN: NDSTR2 OR HSST2.					
OS: Homo sapiens (Human); Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Metazoa; Primates; Catarrhini; Hominidae; Homo.					
OC: NCBI_TaxID:9606;					
RN: [1]					
RP: SEQUENCE FROM N.A.					
RX: MEDLINE=98264822; PubMed=9610156;					
RA: Humphries D.E.; Landolt J.J.; Karlinsky J.B.;					
RT: "cdNA cloning, genomic organization and chromosomal localization of human heparan N-deacetylase/N-sulfotransferase-2."					
RL: Biochem. J. 332:303-307(1998).					
-1- FUNCTION: Catalyzes the N-sulfation and N-deacetylylation of Glucosamine of the Glycosaminoglycan in heparin sulfate.					
CC: Plays a role in determining the extent of modification of the polysaccharide chain.					
-1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. GOLGI MEMBRANE					
CC: (BY SIMILARITY).					
-1- SIMILARITY: BELONGS TO THE SULFOTRANSFERASE FAMILY.					
CC: This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce or send an email to licensee@isb-sib.ch).					
CC: DR: EMBL; U366601; AR27120-1; .					
DR: AP042094; AAB970861; .					
DR: HSSP; P52848; INST.					
DR: Genew; HGNC:7681; NDST2.					
DR: MIM; 603268; .					
DR: GO; GO:0004393; F:heparin N-deacetylase/N-sulfotransferase ac. . . ; TAS.					
DR: InterPro; IPR00863; Sulfotransfere.1.					
DR: Pfam; PF00685; Sulfotransfere.1.					
DR: KN: Transmembrane; Glycoprotein; Golgi stack; Signal-anchor.					
DR: CITOPLASMIC (POTENTIAL).					
DR: DOMAIN; 1 18					
DR: SIGNAL-ANCHOR (TYPB-II MEMBRANE PROTEIN)					
DR: (POTENTIAL).					
DR: LUMINAL, CATALYTIC (POTENTIAL).					
DR: N-LINED (GLCNAC, . . ) (POTENTIAL).					
DR: N-LINED (GLCNAC, . . ) (POTENTIAL).					
DR: N-LINED (GLCNAC, . . ) (POTENTIAL).					
DR: N-LINED (GLCNAC, . . ) (POTENTIAL).					

CC	the European Bioinformatics Institute. There are no restrictions on use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to license@isb-sib.ch).
CC	EMBL; U02304; AAC52137.1; -;
CC	DR EMBL; X75885; CA453479.1; -;
CC	DR FIR; A49733; A4973.
CC	DR HSSP; P52848; INST.
CC	DR MGD; MGI:97040; Nstd2;
CC	DR GO; GO:008151; P; cell growth and/or maintenance; IMP.
CC	DR GO; GO:0015012; P; heparan sulfate proteoglycan biosynthesis; IMP.
CC	DR InterPro; IPR000663; Sulfotransferase.
CC	DR Pfam; PF00685; Sulfotransferase_1.
CC	KW Transferase; Transmembrane; Glycoprotein; Golgi stack; Signal-anchor.
CC	CYTOPLASMIC (POTENTIAL).
CC	FT DOMAIN 1 18 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
CC	FT TRANSMEM 1 19 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
CC	FT DOMAIN 40 883 LUMINAL, CATALYTIC (POTENTIAL).
CC	FT CARBOHYD 233 233 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC	FT CARBOHYD 350 350 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC	FT CARBOHYD 400 400 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC	FT CARBOHYD 666 666 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC	FT CARBOHYD 726 726 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC	FT CARBOHYD 802 802 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC	FT CONFLICT 118 118 A > V (IN REF. 2).
CC	FT CONFLICT 396 396 MISSING (IN REF. 2).
CC	SQ 883 AA; 101202 MW; 8AC8C7BEF5B8EED8 CRC64;
CC	SEQUENCE 883 AA;
CC	Query Match 18.0%; Score 326.5; DB 1;
CC	Best Local Similarity 31.2%; Prd. No. 1.7e-19;
CC	Matches 86; Conservative 48; Mismatches 125; Indels 17; Gaps 7;
CC	CC
CC	QY 77 SKBQRVLHDVQQLPKALLIGVKGGSPTALLEMLNHPAVKA---SQEIHFDNDE 130
CC	DB 595 SKERKT----CDRLPKFLIVPQKGTGTAIHFSLPAVTSSFPSPSTFEETQF-ECP 648
CC	QY 131 NYGRKIEWYRK-KMPFYPQQTIEKSAFVTFEEPVPIRKYNSKILLLIVREPTR 188
CC	DB 649 NYFKGIDYTMDFFPVPSNASTDFLFEEKSYTDFSEVVPREGAILPRAXIMITYLTNPADR 708
CC	QY 306 AGSKGRRIHEPDPSVITKRPKFFPFPNQFYQITGR 341
CC	DB 829 GRSKGRKXPFDMDTIESRLFTTDFFRNENLELSKLSR 864
CC	CC
CC	RESULT 3
CC	HSS1_RAT STANDARD; PRT; 882 AA.
CC	ID HSS1_RAT
CC	AC Q02353;
CC	DT 01-JUL-1993 (Rel. 26, Created)
CC	DT 01-JUL-1993 (Rel. 26, Last sequence update)
CC	DE Heparan sulfate N-deacetylase/N-sulfotransferase (EC 2.8.2.8) (N-HSST)
CC	DE (HSNST) (Heparan sulfate N-sulfotransferase) (N-heparan sulfatase) (Glucosaminyl N-deacetylase/N-sulfotransferase)
CC	DE sulfatase (N-acetylgalactosaminyl N-sulfotransferase)
CC	DE sulfotransferase
CC	GN NDST1 OR HSST1 OR HSST.
CC	OS Rattus norvegicus (Rat).
CC	OC Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi; Teleostei; Actinopterygii; Sarcopterygii; Osteichthyes; Gnathostomata; Tetrapoda; Amniota; Mammalia; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
CC	OX NCBI_TaxID=10116;
CC	RN [1] SEQUENCE FROM N.A., AND SEQUENCE OF 360-375; 568-580; 697-710 AND 813-823.
CC	STRAIN=Leaden X_A1;
CC	MEDLINE=94124588; PubMed=8394489;
CC	Orellana A., Hirschberg C.B., Wei Z., Swidler S.J., Ishihara M.; "Molecular cloning and expression of a glycosaminoglycan N-acetylglucosaminyl N-deacetylase/N-sulfotransferase from a heparin-producing cell line."
CC	J. Biol. Chem. 269:22740-22746 (1994).
CC	[2] SEQUENCE FROM N.A., AND SEQUENCE OF 360-375; 568-580; 697-710 AND 813-823.
CC	STRAIN=Leaden X_A1;
CC	MEDLINE=94193735; PubMed=8144627;
CC	Brikkaeck D., Sandbeck D., Ek B., Lirdahl U., Kjellen L.; "cDNA cloning and sequencing of mouse mastocytoma Glucosaminyl N-deacetylase/N-sulfotransferase, an enzyme involved in the biosynthesis of heparin."
CC	J. Biol. Chem. 269:10438-10443 (1994).
CC	-! FUNCTION: Catalyzes the N-sulfation and N-deacetylation of glucosamine of the glycosaminoglycan in heparin sulfate. Plays a role in determining the extent of modification of the polysaccharide chain.
CC	-! SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. GOLGI MEMBRANE (BY SIMILARITY).
CC	-! SIMILARITY: BELONGS TO THE SULFOTRANSFERASE FAMILY.
CC	This SWISS-PROT entry is copyright. It is produced through collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation -

RESULT 4						
	HSS1	HUMAN	STANDARD;	PRT;	882 AA.	
RC	ID	HSS1	HUMAN			
RA	AC	P52848;				
RA	DT	01-OCT-1996	(Rel. 34, Created)			
RT	DT	01-OCT-1996	(Rel. 34, Last sequence update)			
RT	DT	28-FEB-2003	(Rel. 41, Last annotation update)			
RL	DE	Heparan sulfate N-sulfotransferase (EC 2.8.2.8) (N-HSST)				
RL	DE	(HSST) ([Heparan sulfate]-glucosamine N-sulotransferase) (N-heparan sulfate sulfotransferase) (Glucosaminyl N-deacetylase/N-sulfotransferase).				
RL	GN	NDST1 OR HSST1 OR HSST.				
RT	OS	Homo sapiens (Human).				
RT	OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
RT	OC	NCBI_TaxID=9606;				
RN	RN	[1] -				
RP	RP	SEQUENCE FROM N.A.				
RC	RT	TISSUE=placenta;				
RC	RX	Medline=95324914; PubMed=7601448;				
RC	RA	Dixon J., Loftus S.K., Gladwin A.J., Scambler P.J., Wasmuth J.J., Dixon M.J.;				
RT	RT	"Cloning of the human heparan sulfate-N-deacetylase/N-sulfotransferase gene from the Treacher Collins syndrome candidate region at 5q32-q33.1.";				
RT	RT	Swiss-Prot entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).				
CC	CC	SEQUENCE FROM N.A.				
CC	CC	TISSUE=Connective tissue;				
CC	CC	CYTOSOLIC				
CC	FT	Label T.L., Milewicz D.J., Bonadio J., Edelhoff S., RA				
CC	FT	Genomics 26:239-244 (1995).;				
CC	FT	[2]				
CC	FT	SEQUENCE FROM N.A., AND SUBCELLULAR LOCATION.				
CC	FT	RA				
CC	FT	DISCLOSURE OF HUMAN HEPARAN GLUCOSAMINYL N-DEACETYLASE/N-SULFOTRANSFERASE TO THE TRANS-GOLGI NETWORK."				
CC	FT	RT				
CC	FT	"Localization of human heparan glucosaminyl N-deacetylase/N-sulfotransferase to the trans-Golgi network."				
CC	FT	RA				
CC	FT	MEDLINE=97361922; PubMed=9230113;				
CC	FT	Humphries D.E., Sullivan B.M., Aleixo M.D., Stow J.L.;				
CC	FT	RT				
CC	FT	"Localization of human heparan glucosaminyl N-deacetylase/N-sulfotransferase to the trans-Golgi network."				
CC	FT	RA				
CC	FT	J. Biol. Chem. 325:351-357 (1997).				
CC	FT	[3]				
CC	FT	SEQUENCE FROM N.A.				
CC	FT	RT				
CC	FT	X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS) OF 579-882.				
CC	FT	RA				
CC	FT	RX				
CC	FT	Medline=9914127; PubMed=10196134.				
CC	FT	Kakuta Y., Sueyoshi T., Negishi M., Pedersen L.C., RA				
CC	FT	RT				
CC	FT	"Crystal structure of the sulfotransfase domain of human heparan polysaccharide chain. 3'-phosphoadenosine 3',5'-bisphosphate + [heparan sulfate]-N-sulfoglucosamine."				
CC	FT	-!- CATALYTIC ACTIVITY: 3'-phosphoadenosine 3',5'-bisphosphate + [heparan sulfate]-glucosamine = adenosine 3',5'-bisphosphate + [heparan sulfate]-N-sulfotransferase.				
CC	FT	RT				
CC	FT	"Localizes to the trans-Golgi network."				
CC	FT	J. Biol. Chem. 274:10673-10676 (1999).				
CC	FT	-!- FUNCTION: Catalyzes the N-sulfation and N-deacetylation of Glucosamine of the Glycosaminoglycan in heparan sulfate.				
CC	FT	CC				
CC	FT	PLAYS A ROLE IN DETERMINING THE EXTENT OF MODIFICATION OF THE POLYSACCHARIDE CHAIN.				
CC	FT	RA				
CC	FT	"Crystallization of the sulfotransfase domain of human heparan sulfate N-deacetylase/N-sulfotransfase 1."				
CC	FT	RT				
CC	FT	J. Biol. Chem. 274:10673-10676 (1999).				
CC	FT	-!- SIMILARITY: BELONGS TO THE SULFOTRANSFERASE FAMILY.				
CC	FT	CC				
CC	FT	THIS SWISS-PROT ENTRY IS COPYRIGHT. IT IS PRODUCED THROUGH A COLLABORATION BETWEEN THE SWISS INSTITUTE OF BIOINFORMATICS AND THE EMBL OUTSTATION - THE EUROPEAN BIOINFORMATICS INSTITUTE. THERE ARE NO RESTRICTIONS ON ITS USE BY NON-PROFIT INSTITUTIONS AS LONG AS ITS CONTENT IS IN NO WAY MODIFIED AND THIS STATEMENT IS NOT REMOVED. USAGE BY AND FOR COMMERCIAL ENTITIES REQUIRES A LICENSE AGREEMENT (SEE HTT://WWW.ISB-SIB.CH/ANNOUNCE/ OR SEND AN EMAIL TO LICENSE@ISB-SIB.CH).				
DR	DR	EMBL; U18918; AAA75281.1;				
DR	DR	DR; U36660; AAC27354.1;				
DR	DR	EMBL; U17970; AAA67755.1;				
DR	DR	PIR; U17169; AA5169;				
DR	DR	PDB; INST; 16-SSP-99;				
DR	DR	EMBL; U32530; AAC027354.1;				
DR	DR	EMBL; U17970; AAA67755.1;				
DR	DR	PDB; INST; 16-SSP-99;				
DR	DR	EMBL; U32530; AAC027354.1;				
DR	DR	EMBL; U17970; AAA67755.1;				
DR	DR	PDB; INST; 16-SSP-99;				
DR	DR	EMBL; U32530; AAC027354.1;				
DR	DR	EMBL; U17970; AAA67755.1;				
DR	DR	PDB; INST; 16-SSP-99;				
DR	DR	EMBL; U32530; AAC027354.1;				
DR	DR	EMBL; U17970; AAA67755.1;				
DR	DR	PDB; INST; 16-SSP-99;				
DR	DR	EMBL; U32530; AAC027354.1;				
DR	DR	EMBL; U17970; AAA67755.1;				
DR	DR	PDB; INST; 16-SSP-99;				
DR	DR	EMBL; U32530; AAC027354.1;				
DR	DR	EMBL; U17970; AAA67755.1;				
DR	DR	PDB; INST; 16-SSP-99;				
DR	DR	EMBL; U32530; AAC027354.1;				
DR	DR	EMBL; U17970; AAA67755.1;				
DR	DR	PDB; INST; 16-SSP-99;				
DR	DR	EMBL; U32530; AAC027354.1;				
DR	DR	EMBL; U17970; AAA67755.1;				
DR	DR	PDB; INST; 16-SSP-99;				
DR	DR	EMBL; U32530; AAC027354.1;				
DR	DR	EMBL; U17970; AAA67755.1;				
DR	DR	PDB; INST; 16-SSP-99;				
DR	DR	EMBL; U32530; AAC027354.1;				
DR	DR	EMBL; U17970; AAA67755.1;				
DR	DR	PDB; INST; 16-SSP-99;				
DR	DR	EMBL; U32530; AAC027354.1;				
DR	DR	EMBL; U17970; AAA67755.1;				
DR	DR	PDB; INST; 16-SSP-99;				
DR	DR	EMBL; U32530; AAC027354.1;				
DR	DR	EMBL; U17970; AAA67755.1;				
DR	DR	PDB; INST; 16-SSP-99;				
DR	DR	EMBL; U32530; AAC027354.1;				
DR	DR	EMBL; U17970; AAA67755.1;				
DR	DR	PDB; INST; 16-SSP-99;				
DR	DR	EMBL; U32530; AAC027354.1;				
DR	DR	EMBL; U17970; AAA67755.1;				
DR	DR	PDB; INST; 16-SSP-99;				
DR	DR	EMBL; U32530; AAC027354.1;				
DR	DR	EMBL; U17970; AAA67755.1;				
DR	DR	PDB; INST; 16-SSP-99;				
DR	DR	EMBL; U32530; AAC027354.1;				
DR	DR	EMBL; U17970; AAA67755.1;				
DR	DR	PDB; INST; 16-SSP-99;				
DR	DR	EMBL; U32530; AAC027354.1;				
DR	DR	EMBL; U17970; AAA67755.1;				
DR	DR	PDB; INST; 16-SSP-99;				
DR	DR	EMBL; U32530; AAC027354.1;				
DR	DR	EMBL; U17970; AAA67755.1;				
DR	DR	PDB; INST; 16-SSP-99;				
DR	DR	EMBL; U32530; AAC027354.1;				
DR	DR	EMBL; U17970; AAA67755.1;				
DR	DR	PDB; INST; 16-SSP-99;				
DR	DR	EMBL; U32530; AAC027354.1;				
DR	DR	EMBL; U17970; AAA67755.1;				
DR	DR	PDB; INST; 16-SSP-99;				
DR	DR	EMBL; U32530; AAC027354.1;				
DR	DR	EMBL; U17970; AAA67755.1;				
DR	DR	PDB; INST; 16-SSP-99;				
DR	DR	EMBL; U32530; AAC027354.1;				
DR	DR	EMBL; U17970; AAA67755.1;				
DR	DR	PDB; INST; 16-SSP-99;				
DR	DR	EMBL; U32530; AAC027354.1;				
DR	DR	EMBL; U17970; AAA67755.1;				
DR	DR	PDB; INST; 16-SSP-99;				
DR	DR	EMBL; U32530; AAC027354.1;				
DR	DR	EMBL; U17970; AAA67755.1;				
DR	DR	PDB; INST; 16-SSP-99;				
DR	DR	EMBL; U32530; AAC027354.1;				
DR	DR	EMBL; U17970; AAA67755.1;				
DR	DR	PDB; INST; 16-SSP-99;				
DR	DR	EMBL; U32530; AAC027354.1;				
DR	DR	EMBL; U17970; AAA67755.1;				
DR	DR	PDB; INST; 16-SSP-99;				
DR	DR	EMBL; U32530; AAC027354.1;				
DR	DR	EMBL; U17970; AAA67755.1;				
DR	DR	PDB; INST; 16-SSP-99;				
DR	DR	EMBL; U32530; AAC027354.1;				
DR	DR	EMBL; U17970; AAA67755.1;				
DR	DR	PDB; INST; 16-SSP-99;				
DR	DR	EMBL; U32530; AAC027354.1;				
DR	DR	EMBL; U17970; AAA67755.1;				
DR	DR	PDB; INST; 16-SSP-99;				
DR	DR	EMBL; U32530; AAC027354.1;				
DR	DR	EMBL; U17970; AAA67755.1;				
DR	DR	PDB; INST; 16-SSP-99;				
DR	DR	EMBL; U32530; AAC027354.1;				
DR	DR	EMBL; U17970; AAA67755.1;				
DR	DR	PDB; INST; 16-SSP-99;				
DR	DR	EMBL; U32530; AAC027354.1;				
DR	DR	EMBL; U17970; AAA67755.1;				
DR	DR	PDB; INST; 16-SSP-99;				
DR	DR	EMBL; U32530; AAC027354.1;				
DR	DR	EMBL; U17970; AAA67755.1;				
DR	DR	PDB; INST; 16-SSP-99;				
DR	DR	EMBL; U32530; AAC027354.1;				
DR	DR	EMBL; U17970; AAA67755.1;				
DR	DR	PDB; INST; 16-SSP-99;				
DR	DR	EMBL; U32530; AAC027354.1;				
DR	DR	EMBL; U17970; AAA67755.1;				
DR	DR	PDB; INST; 16-SSP-99;				
DR	DR	EMBL; U32530; AAC027354.1;				
DR	DR	EMBL; U17970; AAA67755.1;				
DR	DR	PDB; INST; 16-SSP-99;				
DR	DR	EMBL; U32530; AAC027354.1;				
DR	DR	EMBL; U17970; AAA67755.1;				
DR	DR	PDB; INST; 16-SSP-99;				
DR	DR	EMBL; U32530; AAC027354.1;				
DR	DR	EMBL; U17970; AAA67755.1;				
DR	DR	PDB; INST; 16-SSP-99;				
DR	DR	EMBL; U32530; AAC027354.1;				
DR	DR	EMBL; U17970; AAA67755.1;				
DR	DR	PDB; INST; 16-SSP-99;				
DR	DR	EMBL; U32530; AAC027354.1;				
DR	DR	EMBL; U17970; AAA67755.1;				
DR	DR	PDB; INST; 16-SSP-99;				
DR	DR	EMBL; U32530; AAC027354.1;				
DR	DR	EMBL; U17970; AAA67755.1;				
DR	DR	PDB; INST; 16-SSP-99;				
DR	DR	EMBL; U32530; AAC027354.1;				
DR	DR	EMBL; U17970; AAA67755.1;				
DR	DR	PDB; INST; 16-SSP-99;				
DR	DR	EMBL; U32530; AAC027354.1;				
DR	DR	EMBL; U17970; AAA67755.1;				
DR	DR	PDB; INST; 16-SSP-99;				
DR	DR	EMBL; U32530; AAC027354.1;				
DR	DR	EMBL; U17970; AAA67755.1;				
DR	DR	PDB; INST; 16-SSP-99;				
DR	DR	EMBL; U32530; AAC027354.1;				
DR	DR	EMBL; U17970; AAA67755.1;				
DR	DR	PDB; INST; 16-SSP-99;				
DR	DR	EMBL; U32530; AAC027354.1;				
DR	DR	EMBL; U17970; AAA67755.1;				
DR	DR	PDB; INST; 16-SSP-99;				
DR	DR	EMBL; U32530; AAC027354.1;				
DR	DR	EMBL; U17970; AAA67755.1;				
DR	DR	PDB; INST; 16-SSP-99;				
DR	DR	EMBL; U32530; AAC027354.1;				
DR	DR	EMBL; U17970; AAA67755.1;				
DR	DR	PDB; INST; 16-SSP-99;				
DR	DR	EMBL; U32530; AAC027354.1;				
DR	DR	EMBL; U17970; AAA67755.1;				
DR	DR	PDB; INST; 16-SSP-99;				
DR	DR	EMBL; U32530; AAC027354.1;				
DR	DR	EMBL; U17970; AAA67755.1;				
DR	DR	PDB; INST; 16-SSP-99;				
DR	DR	EMBL; U32530; AAC027354.1;				
DR	DR	EMBL; U17970; AAA67755.1;				
DR	DR	PDB; INST; 16-SSP-99;				
DR	DR	EMBL; U32530; AAC027354.1;				
DR	DR	EMBL; U17970; AAA67755.1;				
DR	DR	PDB; INST; 16-SSP-99;				
DR	DR	EMBL; U32530; AAC02735				



Db	54	LILSDRILLERACWRYPPHSDEKRVTHVGCKINEPQFQERPSFFA-----ELTAWPQG 104	CC use by non-profit institutions as long as its content is in no way CC modified and this statement is not removed. Usage by and for commercial CC entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> CC or send an email to license@isb-sib.ch).
Qy	176	IKLLIVVREPTRALISYTVQLEGERKRNTRKYKFELAIDP-----NTECEVNTRKIK 227	
Db	105	IKVILVIRNTTLESLSRFVQA----RQTROWLQFQSDSSAAPPPVMLPFATCEA----- 154	
Qy	228	AVRTSITVYKHLB---RWLKYTPPIEQHVVDGRDLITEPLPLQLOLYKEFLNLIP 276	
Db	155	-----YFRAADDHFHARVNAFDSSRTRLIEYERLRLDPVPVCATVLDFLGAP 201	
<b>RESULT 6</b>			
ID	PXA2 YEAST	STANDARD;	EPT; 853 AA.
AC	P34250;		
DT	01-FEB-1994 (Rel. 28, Created)		
DT	01-FEB-1994 (Rel. 28, Last sequence update)		
DT	16-OCT-2001 (Rel. 40, Last annotation update)		
DB	Peroxisomal long-chain fatty acid import protein 1 (Peroxisomal ABC transporter 2).		
DB	GN PA1 OR PA1 OR YKL741.		
OS	Saccharomyces cerevisiae (Baker's yeast).		
OC	Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetidae; Saccharomycetales; Saccharomycetaceae; Saccharomycetes.		
OX	NCBI_TAXID=4932;		
RN	[1]		
RP	SEQUENCE FROM N.A.; PubMed=97030294;		
RX	MBIDLINE=97030294;		
RA	Shani N., Valle D.;		
RT	"A Saccharomyces cerevisiae homolog of the human adrenoleukodystrophy transporter is a heterodimer of two half ATP-binding cassette transporters.";		
RT	Proc. Natl. Acad. Sci. U.S.A. 93:11901-11906(1996).		
RN	[2]		
RP	SEQUENCE FROM N.A.; PubMed=94205264;		
RX	MBIDLINE=94205264;		
RA	Wiemann S., Voss H., Schwager C., Rupp T., Stegemann J.,		
RA	Zimmermann J., Grothues D., Sensen C., Erflie H., Hewitt N.,		
RA	Banerji A., Ansorge W.,		
RA	"Sequencing and analysis of 51.6 kilobases on the left arm of chromosome XI from Saccharomyces cerevisiae reveals 23 open reading frames including the FAS1 gene.",		
RT	Yeast 9:1343-1348(1993).		
RN	[3]		
RP	SEQUENCE FROM N.A.; PubMed=9528165;		
RX	MBIDLINE=9528165;		
RA	Bossier P., Fernandes L., Vilela C., Rodrigues-Pousada C.;		
RA	"The yeast YKL741 gene situated on the left arm of chromosome XI codes for a homologue of the human ALD protein.",		
RT	Yeast 10:681-686(1994).		
RN	[4]		
RP	SEQUENCE FROM N.A.; PubMed=7941751;		
RA	Maia e Silva A., Bossier P., Vilela C., Fernandes L., Soares H.,		
RA	Guerreiro P., Rodrigues-Pousada C.;		
RL	Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.		
RN	[5]		
RP	FUNCTION.		
RX	MBIDLINE=96324389;		
RA	Hettema E.H., van Roermund C.W.T., Distel B., van den Berg M.,		
RA	Vilela C., Rodrigues-Pousada C., Wanders R.J.A., Tabak H.F.;		
RT	"The ABC transporter proteins Pat1 and Pat2 are required for import of long-chain fatty acids into peroxisomes of Saccharomyces cerevisiae.",		
RT	EMBO J. 15:3813-3822(1996).		
CC	- FUNCTION: INVOLVED IN THE IMPORT OF ACTIVATED LONG-CHAIN FATTY ACIDS FROM THE CYTOSOL TO THE PROTOXISOMAL MATRIX.		
CC	- SUBUNIT: Forms an heterodimer with Pat2.		
CC	- SUBCELLULAR LOCATION: Integral membrane protein. Peroxisomal.		
CC	- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MDR SUBFAMILY.		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its		
<b>RESULT 7</b>			
PHS2_RAT	STANDARD;	PRT;	841 AA.
ID	PHS2_RAT		
AC	P09872;		
DT	01-MAR-1989 (Rel. 1.0, Created)		
DT	28-FEB-2003 (Rel. 4.1, Last sequence update)		
DT	15-SEP-2003 (Rel. 4.2, Last annotation update)		
DB	Glycogen phosphorylase, muscle form (EC 2.4.1.1) (Myophosphorylase).		
GN	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.		
NCBI_TAXID=10116;			
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MBIDLINE=9320110;		
RA	Hudson J.W., Hefferton K.L., Cicerar M.M.;		

RT "Comparative analysis of species-independent, isozyme-specific amino-acid substitutions in mammalian muscle, brain and liver glycogen phosphorylases."  
 RT Biochim. Biophys. Acta 1164:197-208 (1993).

[2] RN 565-761 FROM N.A.

RP MEDLINE=86248098; PubMed=2424788;

RA Osawa S., Chiu R.H., McDonough A., Miller T.B. Jr., Johnson G.L.,  
 RT "Isolation of partial cDNAs for rat liver and muscle glycogen  
 phosphorylase isozymes.",  
 RL PDB Lett. 202:282-288 (1986).

[3] RP SEQUENCE OF 762-841 FROM N.A.  
 RX MEDLINE=86030264; PubMed=3840433;

RA Czerar M.M.; See Y.P., Vincentini A.M., Powers M.A., Fletterick R.J.,  
 RA "Comparative sequence analysis of rat, rabbit, and human muscle  
 Glycogen phosphorylase cDNAs.",  
 RT Eur. J. Biochem. 152:267-274 (1985).

-!- FUNCTION: PHOSPHORYLASE IS AN IMPORTANT ALLOSTERIC ENZYME IN  
 CC CARBOHYDRATE METABOLISM. ENZYMES FROM DIFFERENT SOURCES DIFFER IN  
 CC THEIR REGULATORY MECHANISMS AND IN THEIR NATURAL SUBSTRATES.  
 CC HOWEVER, ALL KNOWN PHOSPHORYLASES SHARE CATALYTIC AND STRUCTURAL  
 CC PROPERTIES.

CC -!- CATALYTIC ACTIVITY: {[(1-4)-alpha-D-glucosyl]}(N) + phosphate =  
 CC [(1-4)-alpha-D-glucosyl](N-1) + alpha-D-glucose 1-phosphate.  
 CC -!- SUBUNIT: HOMODIMER. DIMERS ASSOCIATE INTO A TETRAMER TO FORM THE  
 CC ENZYMATIICALLY ACTIVE PHOSPHORYLASE A.

CC -!- PTM: Phosphorylation of Ser-14 converts phosphorylase B  
 CC (unphosphorylated) to phosphorylase A.

CC -!- SIMILARITY: BELONGS TO THE GLYCOCEN PHOSPHORYLASE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC use by non-profit institutions. There are no restrictions on its  
 CC modification and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to license@isb-sib.ch).  
 CC -----

DR EMBL; L10663; AAA1252.1; -.  
 DR EMBL; X03032; CAA26835.1; -.  
 DR PIR; S34624; S34624.

DR HSSP; P0483; IPFG.  
 DR InterPro; IPR000811; Glyco trans\_35.

PFam; PF00343; Phosphorylase; 1\_.  
 DR PROSITE; PS00102; PHOSPHORYLASE; 1.

KW Transferase; Glycosyltransferase; Carbohydrate metabolism;  
 Glycogen metabolism; Allosteric enzyme; Pyridoxal phosphate;  
 Acetylation; Phosphorylation.

RW INIT MET 0 BY SIMILARITY.  
 FT MOD\_RES 1 ACETYLATION (BY SIMILARITY).  
 FT MOD\_RES 14 PHOSPHORYLATION (BY PHK) (IN  
 FT TRANSFERASE; PHOSPHORYLASE A).  
 FT BINDING 75 MAY BE INVOLVED IN AMP BINDING (BY  
 FT SIMILARITY).  
 FT SITE 108 INVOLVED IN THE ASSOCIATION OF SUBUNITS  
 FT (BY SIMILARITY).

FT SITE 142 INVOLVED IN THE ASSOCIATION OF SUBUNITS  
 FT (BY SIMILARITY).

FT ACT\_SITE 155 MAY BE INVOLVED IN THE ALLOSTERIC CONTROL  
 FT OF ENZYME ACTIVITY (BY SIMILARITY).

FT BINDING 680 PYRIDOXAL PHOSPHATE (BY SIMILARITY).  
 FT CONFLICT 639 RF -> L (IN REF. 2).  
 FT CONFLICT 723 Q -> N (IN REF. 2).  
 FT CONFLICT 765 V -> L (IN REF. 3).

SQ SEQUENCE 841 AA; 97141 MW; 80D72C03011A686A CRC64;  
 SQ Score 98.5%; Length 841;  
 Best Local Similarity 5.4%; Pred. No. 2.2%;  
 Matches 66; Conservative 39; Mismatches 92; Indels 115; Gaps 18;

Qy 123 INFDDNDENYKGIELWTRKK---MFPSYP-----QITIEKSPAYFITEPVPE 167

Db 200 VHFYGRVEHTSGAKWVDTQVLYAMVDTPPGYNNVVNTMRLLGAKAPPYENLKD--- 256

Qy 168 RIYKANSSIKILLI--IVREPTPTRAISDYTOVLEGERKNTYY---- 208

Db 257 -FNGGGIQAVLDRNLAENISRVLVNPDKFEGNEKLROEYFVVAATLQDIIRFKSS 314

Qy 209 KF-----EKLAIDPNTCEVNNTKYK----AVTSIX 234

Db 315 KFGCRDPVRNTEDAFFDKVQLNDHPSLAIPELIRLYDWERLWDVTVTCYAC 374

Qy 235 TKH----LERWLKYPFQFHVVDRLITEPL-EQLV---EKFLN----LPPR 278

Db 375 TNHTVLPPEALRW----PV-----HEMTEFLPRHLQIYEVNORFLNRVAAAFPGD 421

Qy 279 ISQNNYFNATRGFYCLRFNTIFNCLAGSK-----GRTHPEVDPVTILRKFF---H 329

Db 422 VDLRMSLVEBG-AVKRINNA-HLCAGSHAVINGVARISEILKKTIFK--DFVBLEPH 477

RESULT 8  
 PAP1\_POWPV STANDARD; PRT; 472 AA.

ID PAP1\_POWPV  
 AC Q9J5BB;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)

DB Poly(A) polymerase catalytic subunit (EC 2.7.7.19) (Poly(A) polymerase  
 DE large subunit) (PAP-L) (WP55).  
 GN PPV102.

OS Fowlpox virus (FPV).

Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;

OC Avipoxvirus.

CC NCBI\_TaxID=10211;  
 RN [1] -----

RP SEQUENCE FROM N.A.  
 RX MEDLINE=20193830; PubMed=10729156;

RA Afonso C.L., Tulman E.R., Lu Z., Zeak L., Kutish G.F., Rock D.L.;  
 RT "The genome of fowlpox virus." ;  
 RL J. Virol. 74:315-331 (2000).

CC -!- FUNCTION: POLYMERASE THAT CREATES THE 3' POLY(A) TAIL OF MRNA'S.  
 CC -!- CATALYTIC ACTIVITY: N ATP + [nucleotide] (M+N) = N diphosphate +  
 CC (nucleotide) (M+N).  
 CC -!- SUBUNIT: HETERODIMER OF VP55 (CATALYTIC) AND VP39 (REGULATOR).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to license@isb-sib.ch).  
 CC -----

DR EMBL; AF198100; AAF44446.1; -;  
 DR IntePro; IPR04976; Pox\_PolyA\_pol.  
 DR Pfam; PF03296; Pox\_PolyA\_Pol\_1.  
 KW mRNA Processing; Transferase; Transcription.  
 FT ACT\_SITE 194 BY SIMILARITY.  
 FT ACT\_SITE 196 BY SIMILARITY.  
 SQ SEQUENCE 472 AA; 56081 MW; ADDD6F3RCB8AD022 CRC64;  
 SQ Score 97.5%; Length 472;

Qy 67 LIHFEPKGNASKE---QVRLHDIVQCLPAIIIGVRKGTRALLEMINLH-PAVVKASQ 121

Db 134 ILDSNVVPEKSVKGRHKVSDLVHVVYK-----IMEYLRRNSNSLCYGSY 181

QY	122 EIHFFDNDENYG-----KGIEIYRKOMP-----	145	DR GO:0000742; P:karyogamy during conjugation with cellular . . ; IGI.
Db	182 SLHFUNNKVKEGIDYLQTNARTFLINIAFLIKEITGRRLVILKPYLQVYIMHEETIN	241	DR GO:0006457; P:protein folding; DnaJ_N.
QY	146 SYPQQTIEKSPAYFTEEVPERIYKMNNSIKLLIVTREPTTRALSDYTQLEGERKDKK	205	DR IntPro; IPR001623; DnaJ_1.
Db	242 HVMDTFNIRKETNMMPKLMIDNMIVDPCIQQLNNIK-----MLSQIDRLEE	289	DR SMART; SM00271; DnaJ_1.
QY	206 TYKKEPKLAIDNTCEHVNTKYK-----AVRTSI-----YTKHLERWLK	243	DR PROSITE; PS00036; DnaJ_1; FALSE_NEG.
Db	290 LQAKEKLSSLRGTLLEYTRYRYSIPLDSSILEVRAKLDKDKRKITVDEKKYKVNHYRC	349	DR PROSITE; PS50076; DnaJ_2; 1.
QY	244 YFPIEQHVVGDRLITEPLPFLQVLEPKVNLPPRISQVNLYENATRGFCYCLR----ENI	299	KW Hypothetical protein.
Db	350 YFYLDEVE-----LKKFPSKNSGEDEYDFEAVTNSEYFAIRNKTMTY	392	FT DOMAIN 560 569 POLY_GLN.
QY	300 IENKCLAGSKGRHIPEVDPSVITKLRKFEPFHNPQKY	336	FT DOMAIN 585 655 J-DOMAIN.
Db	393 FSNTALMRSENEIHP-ITINWLTSHALLYTTRKEY	428	SQ SEQUENCE 632 AA; 80381 MW; 9F612DD16B66981B CRC64;
QY	351 -----	1	Query Match 5.4%; Score 97.5%; DB 1; Length 692;
Db	313 ISKUNKVNSPSKRQI-----	1	Best Local Similarity 20.2%; Pred. No. 2.1;
QY	336 NKEFRSWDRLTIEPFKDKKEPFITMKLINKDTNFKNYVFFELIKQLEEDVQLS-	1	Matches 62; Conservative 48; Mismatches 90; Indels 107; Gaps 15;
Db	313 ISKUNKVNSPSKRQI-----	1	SDPAT-----YAYFE 335
QY	68 LHEFRKGNAKSEKQVRLHDLVQQLPKAIIIGVRKGCCRALLEMLNHPAVVKASQBIIHFFD	127	68 LHEFRKGNAKSEKQVRLHDLVQQLPKAIIIGVRKGCCRALLEMLNHPAVVKASQBIIHFFD 127
Db	313 ISKUNKVNSPSKRQI-----	1	Query Match 5.4%; Score 97.5%; DB 1; Length 692;
QY	128 NDE-NYGCIEIWK-----KMPESYFPOQI-----TIEKSPAYFITEEUPERIYKANSSIKLLI	180	Best Local Similarity 20.2%; Pred. No. 2.1;
Db	336 NKEFRSWDRLTIEPFKDKKEPFITMKLINKDTNFKNYVFFELIKQLEEDVQLS-	1	Matches 62; Conservative 48; Mismatches 90; Indels 107; Gaps 15;
QY	336 NKEFRSWDRLTIEPFKDKKEPFITMKLINKDTNFKNYVFFELIKQLEEDVQLS-	1	SDPAT-----YAYFE 335
Db	313 ISKUNKVNSPSKRQI-----	1	Query Match 5.4%; Score 97.5%; DB 1; Length 692;
QY	181 IVREPTTRALSDYTQVLEGERKKNYTYKFEKLA-1DPNTCEVNTK-----	225	Best Local Similarity 20.2%; Pred. No. 2.1;
Db	391 --PFLAKNLIFEDPPTDFV-KPKSYHTTDLYKIDSILCQASSMSPDYKRAKLAAPC	446	Matches 62; Conservative 48; Mismatches 90; Indels 107; Gaps 15;
QY	226 YKAYRTSTYTKHLERWLKYPFIEQFHVVGDRLITEPLPBLQIVYEKFNLNPTRISQYNY	285	SDPAT-----YAYFE 335
Db	447 KGSLSRHSL--TLTWTWKH-----ODAKSQQPLPETVLSDW-NSNPHLMMYVW-	492	Query Match 5.4%; Score 97.5%; DB 1; Length 692;
QY	286 FNATRGFCYCLRFNTIINKLAGSKGRHSEVDPVSITYTURKEF-----HPEFNQKRY	336	Best Local Similarity 20.2%; Pred. No. 2.1;
Db	493 -----NSTLINK---SRSKPSQSFKQGLDQINKEFQDNGLSESTNPYVMKQP	536	Matches 62; Conservative 48; Mismatches 90; Indels 107; Gaps 15;
QY	337 QITGRTL 343	1	SDPAT-----YAYFE 335
Db	537 RILQQQL 543	1	Query Match 5.4%; Score 97.5%; DB 1; Length 692;
<hr/>			
RESULT 10			
NCB	NODB RHIME STANDARD; PRT; PRT;	1	NCB ID _NODB RHIME STANDARD; PRT; PRT; 247 AA.
AC	AC P06236; AC P06236;	1	AC P06236; AC P06236;
DT	DT 01-JAN-1988 (Rel. 0.6, Created)	1	DT 01-JAN-1988 (Rel. 0.6, Created)
DT	DT 01-NOV-1995 (Rel. 31, Last sequence update)	1	DT 01-JAN-1988 (Rel. 0.6, Last sequence update)
DT	DT 28-FEB-2003 (Rel. 41, Last annotation update)	1	DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE	DE Hypothetical 80.4 kDa protein in SMC3-MRP18 intergenic region.	1	DE Nodulation protein H (EC 2.8.2.-) (Host-specificity of nodulation protein D).
GN	GN YJL033W OR J1083 OR HRC558.	1	DE NODH OR HSND OR RA0464 OR SMA0851.
OS	OS Saccharomyces cerevisiae (Baker's yeast).	1	GN Rhizobium meliloti (Sinorhizobium meliloti).
OC	OC Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;	1	OS Plasmid pSymA (megaplasmid 1).
OC	OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.	1	OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales; Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.
RN	RN [1] NCBI_TaxID=382; NCBI_TaxID=383;	1	OC NCBI_TaxID=382; NCBI_TaxID=383;
RP	RP SBQDNAE FROM N.A.	1	RN SBQDNAE FROM N.A.
RC	RC STRAIN=RCR2011 / SU47;	1	RC STRAIN=RCR2011 / SU47;
RC	RC MEDLINE=07016382; PubMed=3020515;	1	RC MEDLINE=07016382; PubMed=3020515;
RX	RX Fisher R.F., Swanson J.A., Mulligan J.T., Long S.R.;	1	RX Fisher R.F., Swanson J.A., Mulligan J.T., Long S.R.;
RA	RA "Extented region of nodulation genes in Rhizobium meliloti 1021. II. Nucleotide sequence, transcription start sites and protein products."	1	RA "Extented region of nodulation genes in Rhizobium meliloti 1021. II. Nucleotide sequence, transcription start sites and protein products."
RA	RA Genetics 117:191-201(1987).	1	RA Genetics 117:191-201(1987).
RT	RT [3]	1	RT [3]
RT	RT This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the ENBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).	1	RT This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the ENBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC	CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the ENBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).	1	CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the ENBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC	CC SEQUENCE FROM N.A.	1	CC SEQUENCE FROM N.A.
RC	RC Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.	1	RC Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
RN	RN SEQUENCE OF 135-692 FROM N.A.	1	RN SEQUENCE OF 135-692 FROM N.A.
RN	RN STRAIN=S88BC;	1	RN STRAIN=S88BC;
RN	RN MEDLINE=95282514; PubMed=7762302;	1	RN MEDLINE=95282514; PubMed=7762302;
RX	RX Vandenobo M., Durand P., Dion C., Portetelle D., Hilger F.:	1	RX Vandenobo M., Durand P., Dion C., Portetelle D., Hilger F.:
RA	RA "Sequence of a 1.17 kb DNA fragment from chromosome X of Saccharomyces cerevisiae includes the mitochondrial ribosomal protein L8." Yeast 11:57-60(1995).	1	RA "Sequence of a 1.17 kb DNA fragment from chromosome X of Saccharomyces cerevisiae includes the mitochondrial ribosomal protein L8." Yeast 11:57-60(1995).
RT	RT SIMILARITY: Contains 1 J domain.	1	RT SIMILARITY: Contains 1 J domain.
CC	CC -!- SIMILARITY: TO C.ELEGANS F22B7.5.	1	CC -!- SIMILARITY: TO C.ELEGANS F22B7.5.
CC	CC RN	1	CC RN
DR	DR SBQDNAE FROM N.A.	1	DR SBQDNAE FROM N.A.
DR	DR STRAIN=1021;	1	DR STRAIN=1021;
DR	DR Fisher R.F., Swanson J.A., Mulligan J.T., Long S.R.;	1	DR Fisher R.F., Swanson J.A., Mulligan J.T., Long S.R.;
DR	DR "Extented region of nodulation genes in Rhizobium meliloti 1021. II. Nucleotide sequence, transcription start sites and protein products."	1	DR "Extented region of nodulation genes in Rhizobium meliloti 1021. II. Nucleotide sequence, transcription start sites and protein products."
DR	DR RN	1	DR RN
DR	DR SEQUENCE FROM N.A.	1	DR SEQUENCE FROM N.A.
DR	DR Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.	1	DR Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
EMBL	EMBL X88851; CAA61312; -.	1	EMBL X88851; CAA61312; -.
EMBL	EMBL Z34288; CAA84049.1; -.	1	EMBL Z34288; CAA84049.1; -.
EMBL	EMBL P08632; 1BOZ.	1	EMBL P08632; 1BOZ.
EMBL	EMBL SGD; S0005609; JEM1.	1	EMBL SGD; S0005609; JEM1.
EMBL	EMBL C: endoplasmic reticulum; IDA.	1	EMBL C: endoplasmic reticulum; IDA.
DR	DR GO; GO:0000300; C: peripheral membrane protein of membrane fra. . . ; IDA.	1	DR GO; GO:0000300; C: peripheral membrane protein of membrane fra. . . ; IDA.
DR	DR GO; GO:0003767; F: co-chaperone activity; IGI.	1	DR GO; GO:0003767; F: co-chaperone activity; IGI.
RP	RP SEQUENCE FROM N.A.	1	RP SEQUENCE FROM N.A.

RC	STRAIN=1021;	RA	Dufort I., Tremblay Y., Belanger A., Labrie F., Luu-The V.,
RX	MEDLINE=11481432;	RA	"Isolation and characterization of a stereospecific
RX	Barnette M.J., Fisher R.F., Jones T., Komp C., Abola A.P.,	RT	3beta-hydroxysteroid sulfotransferase (pregnenolone sulfotransferase)
RA	Barley-Hubler F., Bowser L., Capela D., Galibert F., Goury J.,	RT	CDNA."
RA	Gurjal M., Hong A., Ruiz L., Hyman R.W., Kahn D., Kahn M.L.,	RL	DNA Cell Biol. 15:481-487 (1996).
RA	Kalman S., Keating D.H., Palm C., Peck M.C., Surzycki R., Wells D.H.,	RN	[21]
RA	Yeh K.-C., Davis R.W., Fedderspiel N.A., Long S.R.,	RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RT	"Nucleotide sequence and predicted functions of the entire	RT	STRAIN=NTH 2; TISSUE=Adrenal Gland;
RT	sphingolipid meiolitin PSMA megaplasmid."	RC	MEDLINE=26125350; PubMed=8554560;
RL	Proc. Natl. Acad. Sci. U.S.A. 98:9883-9888(2001).	RA	Luu N.X., Driscoll W.J., Martin B.M., Strott C.A.,
CC	-!- FUNCTION: REQUIRED FOR THE FORMATION OF SULFATED NOD FACTOR.	RT	"Molecular cloning and expression of a guinea pig 3-hydroxysteroid sulfotransferase distinct from chiral-specific 3 alpha-hydroxysteroid sulfotransferase."
CC	PROPOSED TO TRANSFER ACTIVATED SULFATE (PAPS) TO A N-	RT	sulfotransferase."
CC	ACETYLGLUCOSAMINE OF THE NOD FACTOR.	RL	Biochem Biophys Res Commun. 217:1078-1086(1995).
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration	CC	FUNCTION: Catalyzes the sulfation of 3-beta-hydroxy1 groups of neutral steroids. High preference for C21 steroid (pregnenolone).
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -	CC	-!- CATALYTIC ACTIVITY: 3'-phosphoadenosylsulfate + an alcohol = adenosine 3',5'-bisphosphate + an alkyl sulfate.
CC	the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to license@isb-sib.ch).	CC	-!- SUBUNIT: Homodimer (By similarity).
CC	CC -!- TISSUE SPECIFICITY: LIVER, INTESTINE AND KIDNEY.	CC	-!- SUBCELLULAR LOCATION: Cytoplasmic.
CC	CC -!- SIMILARITY: BELONGS TO THE SULFOTRANSFERASE FAMILY.	CC	-!- SIMILARITY: LIVER, INTESTINE AND KIDNEY.
DR	EMBL; X04380; CAA27963_1; -.	CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to license@isb-sib.ch).
DR	EMBL; M37411; AAA26339_1; -.	DR	CC
DR	EMBL; AE007237; AAQ65122_1; -.	DR	EMBL; U55944; AAB07868_1; -.
DR	PIR; D24706; D24706.	DR	EMBL; U55115; AAC2347_1; -.
DR	PIR; H95319; H95319.	DR	PTR; JC4531; JC4531.
DR	InterPro; IPR00863; Sulfotransferase.	DR	HSSP; P5024; 1.CM.
DR	Pfam; PF00685; Sulfotransferase_1.	DR	InterPro; IPR00863; Sulfotransferase.
DR	Nodulation; Transferase; Plasmid; Complete proteome.	DR	Pfam; PF07218; Sulfotransferase_1.
KW	DOMAIN 1 17 HYDROPHOBIC.	DR	ProDom; PD0024; 1.CM.
FT	SEQUENCE 247 AA; 28584 MW; 545BC0517BAPEBD3 CRC64;	DR	Transferase; Steroid metabolism.
SQ	Best Local Similarity 5.3%; Score 97; DB 1; Length 247;	KW	INITIATOR 0 BY SIMILARITY.
Matches 42; Conservative 48; Mismatches 69; Indels 74; Gaps 10;	FT	BINDING SITE (POTENTIAL).	
84 HDLVQQLPKAIIIGVKGTRALLEMINHPPAVKKAQSEIHFFPDNDENYGGIIEWYRK - 141	FT	CONFLICT 204 254 PAPS BINDING SITE (POTENTIAL).	
3 HSTLPPQPEFA-TIAMPTGTHYLEELVNEHPNVLSNGELLNTDTN -----WPDLER 53	FT	SEQUENCE 286 AA; 23B83665016CF5A4B CRC64;	
142 -----KMPPSYQQIT-----IEKSPAYTIEPYPERITYKNS -S 175	FT	Query Match 5.3%; Score 97; DB 1; Length 286;	
54 LLLSDRELLERAFLRYPHDSDKRVTHVGCKINPQFQBRSPFFA-----ELTAWPG 104	FT	Best Local Similarity 21.0%; Pred. No. 0.69;	
Qy	176 IKLLITVYRPTTRAIISDYTQVLEGKERKNTKYKFEKIALDP-----NTCEVNNTYK 227	FT	Matches 65; N mismatches 42; N mismatches 114; Indels 88; Gaps 13;
Db	105 LKVLLVRRNTLESLSRPTQA----RQTROWLKEFSDSSAPPVVMPLPATCEA---- 154	Qy	56 FPIGRALQFRGLHLHEFRKGNASKEQYRLDVOQLPKAIIYGVKGCTRALEMLNLH 113
Qy	228 AVRSTSYTKHLE---RWLKYFPIEOFQHVVDGDRLLTPELQLVKEFLNLP 276	Db	12 FPM--VGFSPBLKREVRDFKLVDK-----DTIWTYVKSGTNWNLIEIVLILS 58
Db	155 -----YKRAADDHFHARVVAFDSSIRLIEYRLRDPVPCVATVLDGAP 201	Qy	114 PAVVKAQSOEHTEDDN----DENYGRGKIEWYRKMKMFSPYQQUITIEKSPPAYTIEEYPERI 169
Db	RESULT 11 SUHB_CAVPO STANDARD; PRT; 286 AA.	Db	59 KGDDKWKVQSVPIWDRSPWIETOHGENL-MRSQDKDPIRYTSHLPLHFKSF----- 108
ID	52871; DT 01-OCT-1996 (Rel. 34, Created)	Qy	170 YKANSSSIKLLIVIYREPTPRAISDVTQVLEGERKERKNTKYKFP -KLAIDPNTCEVNNTYKKA 228
AC	DT 01-OCT-1996 (Rel. 34, Last sequence update)	Db	109 --FSSSKAKVYCLRNPRDVLYSGY-----YFTSKMKAIEKPELQQYMKWF 153
DT	15-SEP-2003 (Rel. 42, Last annotation update)	Qy	229 VRTSYY---TKHLERWLKYFPIEOFQHVVDGDRLLTPELQLVKEFLNLP-----
DB	3-beta-hydroxysteroid sulfotransferase (EC 2.8.2.2) (Alcohol	Db	154 QGNTVYGSWEHYRDWSMREKENPJLVSYEELIKDTRSTVEKICQFLGKRLKPEEIDLV 213
DE	sulfotransferase) (Pregnenolone sulfotransferase 2) (HST2).	GN	STD2.
OS	Cavia porcellus (Guinea pig).	Qy	278 -----RISQONLYFN--ATRGFYCLR-----FNLIFNKCL 305
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Buteleostomi;	Db	214 LKTSSEFRKRENENSNYSJLPPNDTTEGFTFLRKGVGDWKHHTVQAQABEFDKTYQERM 273
OC	Mammalia; Buteraria; Rodentia; Hystricognathi; Caviidae; Cavia.	Qy	306 AGSKGRTHP 314
OC	NCBI_TaxID10141;	Db	274 AGYPPKLF 282
RC	SEQUENCE FROM N.A.		
RC	TISSUE=Blood		
RC	MEDLINE=9566475; PubMed=9572014;		

RESULT 12		RESULT 13	
SURA_BUCAI	STANDARD;	AGO1_ARATH	STANDARD;
ID_P57240;	PRT; 430 AA.	ID_004379;	PRT; 1048 AA.
AC DT 16-OCT-2001 (Rel. 40, Created)		DT 16-OCT-2001 (Rel. 40, Created)	
DT 16-OCT-2001 (Rel. 40, Last sequence update)		DT 16-OCT-2001 (Rel. 40, Last sequence update)	
DT 16-OCT-2001 (Rel. 40, Last annotation update)		DT 28-FEB-2003 (Rel. 41, Last annotation update)	
DE Survival protein sura homolog precursor.		DE Argonaute protein.	
GN SURA OR BU140.		GN AGO1 OR AVIG810 OR FILA17.3.	
OS Buchnera aphidicola (subsp. <i>Acyrthosiphon pisum</i> ) (Acyrthosiphon pisum symbiotic bacterium)		OS Arabidopsis thaliana (Mouse-ear cress).	
OS Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Buchnera.		OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.	
OC NCBI_TaxID=118039;		OC NCBI_TaxID=3702;	
OX RN		OX RN	
RP SEQUENCE FROM N.A.		RP SEQUENCE FROM N.A.	
RC STRAIN=TOKYO 1998;		RC STRAIN=cv_Columbia; TISSUE=Leaf;	
RX MEDLINE=20445123; PubMed=10993077;		RX MEDLINE=98090460; PubMed=9427751;	
RA Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;		RA Bohmert K., Camus I., Bellini C., Bouchez D., Caboche M., Benning C.,	
RT "Genome sequence of the endocellular bacterial symbiont of aphids Buchnera sp. Aps."		RT "AGO1 defines a novel locus of Arabidopsis controlling leaf development."	
RL Nature 407:81-86 (2000).		RL EMBO J. 17:170-180 (1998).	
CC -!- SIMILARITY: BELONGS TO THE PPIC/PARTYLIN FAMILY OF ROTAMASES.		RN [1]	
CC SEQUENCE FROM N.A.		RN [2]	
CC STRAIN=cv_Columbia;		RN [3]	
RC MEDLINE=21016719; PubMed=11110712;		RC SEQUENCE FROM N.A.	
RX THELOGIS A., ECKER J.R., PALM C.J., FEDERER-SPIEL N.A., KAUL S., RA White O., ALONSO J., ALTAIFI H., BOWMAN C.L., BROOKS S.Y., RA BUEHLER E., CHAN A., CHAO Q., CHEN H., CHEN R.F., CHIN C.W., RA CHUNG M.K., CONN L., CORWAY A.R., DEWAR K., RA DUNN P., ETGU P., FELDBYUM T.V., FENG J.-D., FONG B., FUJII C.Y., RA GILL J.E., GOLDSMITH A.D., HAAS B., HANSEN N.F., HUGHES B., HUIZAR L., RA HUNTER J.L., JANKINS J., JOHNSON-HOPSON C., KURTZ D.B., KUHN S., KRAYKIN E., RA KIM C.J., KOO H.L., KREMENETSKAIA I., KURTZ D.B., KWAN A., LAM B., RA LANGIN-HOOPER S., LEE A., LEE J.M., LENZ C.A., LI J.H., LI Y.-P., RA LIN X., LIU Z.A., LIUOS R., LIUTRI R., MARZILLI A., RA MILLITSCHER J., MIRANDA M., NGUYEN M., NIERMANN W.C., OSBORNE B.I., RA PAI G., PETERSON J., PHAM P.K., RIZZO M., ROONEY T., ROWLEY D., RA SAKANO H., SALZBERG S.L., SCHWARTZ J.R., TAMBUNGA G., SHINN P., SOUTHWICK A.M., RA SAN H., TALLON L.J., TAMBUNGA G., TORIUMI M.J., TOWN C.D., RA UTTERBACK T., VAN AKEN S., VAYSBERG M., VYSOTSKAIA V.S., WALKER M., RA WU D., YU G., FRASER C.M., VENTER J.C., DAVIS R.W.; RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis thaliana";	RA NATURE 408:816-820(2000).		
CC DR InterPro; IPR00297; Rotamase.		CC -!- FUNCTION: ESSENTIAL FOR PROPER DEVELOPMENT OF LEAVES AND FLORAL	
CC PFAM; PF00639; Rotamase; 1.		CC ORGANs, AND FORMATION OF AXILLARY MERISTEMS.	
DR PROSITE; PS01036; PPIC_PPASE_1; FALSE_NEG.		CC -!- SUBCELLULAR LOCATION: Cyttoplasmic (Potential).	
DR PROSITE; PS50198; PPIC_PPASE_2; 2.		CC -!- TISSUE SPECIFICITY: WIDELY EXPRESSED AT LOW LEVELS.	
KW Isomerase; Rotamase; Repeat; Signal; Complete proteome.		CC -!- DEVELOPMENTAL STAGE: EXPRESSED THROUGHOUT ALL DEVELOPMENTAL	
FT SIGNAL 1 20 POTENTIAL.		CC -!- STAGES.	
FT CHAIN 21 430 SURVIVAL PROTEIN SURA HOMOLOG.		CC -!- SIMILARITY: BELONGS TO THE ARGONAUTE FAMILY.	
FT DOMAIN 179 277 PPIC 1.		CC -!- SIMILARITY: Contains 1 PAZ domain.	
FT DOMAIN 286 386 PPIC 2.		CC -!- SIMILARITY: Contains 1 Piwi domain.	
SQ SEQUENCE 430 AA; 50737 MW; CA4C425B3FB45788 CRC64;		CC -!- SIMILARITY: Contains 1 Piwi domain.	
Query Match Score 97; DB 1; Length 430;		CC This SWISS-PROT entry is copyright. It is produced through a collaboration	
Best Local Similarity 20.7%; Pred. No. 1.2;		CC between the Swiss Institute of Bioinformatics and the EMBL Outstation	
Matches 73; Conservative 56; Mismatches 117; Indels 106; Gaps 15;		CC in the European Bioinformatics Institute. There are no restrictions on its	
QY 23 SLLYLVARGSGLDRLOQICPLEGRIGG----ARTQAEE--PLRALQFKRGLLHEP 71		CC use by non-profit institutions as long as its content is in no way	
Db 14 SIFYVLAKNQYDNTRTAIVDETTLNANSVNEILVFLKKKTFPLKSDFLKVKLEU 73		CC modified and this statement is not removed. Usage by commercial	
Qy 72 -----RKGNAKSY-----RLHDIVQQLKATIGVRKGTRALLEML 110		CC entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>	
Db 74 IVDSLILQEANSKNINITEQIDTVIKNIALKKHISVDHFKKQILIRNIK- 123		CC or send an email to license@isb-sib.ch).	
Qy 111 NLHPAVVKASOBHFFDDNDENYGGKGLEWYRK--MFPSYQQTLIEKSPAYFITEEV- 165		CC EMBL; U91995; AAC18440_1; -.	
Db 124 -NPS -----YD---NFKCKBILKMTQDYEHKRNTISEQEVNTLFKKLIK 169		DR EMBL; AC007932; AAD4975_1; -.	
Qy 166 -PERIXKMNSSIKLILIVREPITRALSITQVLEGKERKNTYKPEKLADPNTCVNT 224		DR InterPro; IPR03100; PAZ;	
Db 170 DNEKFCKTINNSYLPKQSDNAYRNTRKIABNIVYKKKGTYFEKJLIE--CEKIK 226		DR InterPro; IPR03115; PAZ;	
Qy 225 KYKAVRTSIYTKHBLWKYRPIEQFHWDRLITEPLPELVYKEFNLPPRSQYNL 284		DR Pfam; PF02170; PAZ; 1.	
Db 227 STIVKMKFWPKLDDQNSSF-----KTLNIFKRGQLGPIGD--- 265		DR PROSITE; PS50821; PAZ; 1.	
Qy 285 YENATRGFYCLERENIENKNGKSKGR-----HPEVDSV-----TKRKR 326		DR PROSITE; PS50822; PAWI; 1.	
Db 266 -----KGLYIILKVNDFHKK-----KENIVTEFYMQHCLIKTESVLTNTTEAKK 307		KW Developmental Protein.	
		DOMAIN 391 501	
		FT	

FT DOMAIN	676	997	PIWI.	
FT DOMAIN	13	104	GLY-RICH.	
SQ SEQUENCE	1048 AA;	116190 MN;	3EE146343A090541 CRC64;	
Query Match Score	97;	DB 1;	Length 1048;	
Best Local Similarity	5.3%;	Pred. No. 4;		
Matches	22.3%;	Mismatches	97;	Indels 13;
QY	10 ROKLILVGSLAVGSLVYLVARYGSLLRQLQPCIEGLGARTQAEFLP-----RALQF 63			
Db	251 RKSLYTAPGPPLPNSEK--RNLNLD-----EVAGGOREREFKVVKLVARADLH 300			
QY	64 KRGLLHEFRKGNAKEDQVRLHDLY-QOLPKALLIGVYRKGGTRALLEMINLHPAVVKASQE 122			
Db	301 HLGMFLEGKQSDAPOSALQVQDVLRELPTEVRYPIVGRS-----339			
QY	123 IFFFDND---ENYKGKIE ---WYRKMPSPSYQQOTIEKSPAYITEVEPTYKNS 174			
Db	340 --FYSPPDGKQSLGDGELESWRGFYQSIPTQMGSLINIDMSSTA-----EANP 388			
QY	175 SIKLL---IYREPLPTTAISDTQVLESKERKNKTYKTFKEKLADPNTCENITKYK-AVRL 231			
Db	389 VIQFQCDLNRDISSRPLSDADRV----KIKKALRGKVIBV-THRGNMRKRYLSGLT 441			
QY	232 STYTKH----ERWKYKPFLEOFHYDGDRLITEPLPELQL 268			
Db	442 AVATRELTFPVDERNTOKSVVYEHETYGGFRITQHTOLPCQLQV 483			
RESULT 14 GYRB_BUCAPI ID_GYRB_BUCAPI	STANDARD;	PRT;	803 AA.	
AC P2945;				
DT 01-APR-1993 (Rel. 25, Created)				
DT 28-FEB-2003 (Rel. 41, Last sequence update)				
DT 28-FEB-2003 (Rel. 41, Last annotation update)				
DE DNA Gyrase subunit B (EC 5.99.1.3).				
GN GYRB OR BUSCO10.				
OS Buchnera aphidicola (subsp. Schizaphis graminum).				
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;				
OC Enterobacteriaceae; Buchnera.				
OX NCBI_TaxID:98794;				
RN [1]				
RP SEQUENCE FROM N.A.				
RX MEDLINE=984963; PubMed=151654;				
RA Clark M.A., Baumann L., Baumann P.;				
RT "Sequence analysis of a 34.7-kb DNA segment from the genome of Buchnera aphidicola (endosymbiont of aphids) containing groEL, dnaA, the atp operon, gida, and rho.";				
RT Curr. Microbiol. 36:158-163 (1998).				
RN [2]				
RP SEQUENCE FROM N.A.				
RX MEDLINE=208459; PubMed=12089438;				
RA Tamas J., Klasson L., Canbäck B., Naesslund A.K., Eriksson A.-S., Werngren J.J., Sandstrom J.P., Moran N.A., Andersson S.G.E.;				
RT "50 million years of genomic stasis in endosymbiotic bacteria." Science 296:2379-2379 (2002).				
RN [3]				
RP SEQUENCE OF 1-91 FROM N.A.				
RX MEDLINE=92241666; PubMed=1572539;				
RA Lai C.-Y., Baumann P.;				
RT "Genetic analysis of an aphid endosymbiont DNA fragment homologous to the rnpA-rnpH-dnaA-dnaN-gtrB region of <i>Escherichia coli</i> ." Gene 113:175-181 (1992).				
CC -!- FUNCTION: DNA gyrase negatively supercoils closed circular double-stranded DNA in an ATP-dependent manner and also catalyzes the interconversion of other topological isomers of double-stranded DNA rings, including catenanes and knotted rings.				
CC -!- CATALYTIC ACTIVITY: ATP-dependent breakage, passage and rejoining of double-stranded DNA.				
CC -!- SUBUNIT: MADE UP OF TWO CHAINS. THE A CHAIN IS RESPONSIBLE FOR DNA BREAKAGE AND REJOINING; THE B CHAIN CATALYZES ATP HYDROLYSIS. THE ENZYME FORMS AN A2B2 TETRAMER.				
RESULT 15 RPOD_SPOIOL				
ID_POD_SPOIOL				STANDARD;
AC P11704;				PRT; 1361 AA.
DT 01-OCT-1989 (Rel. 12, Created)				
DT 01-OCT-1989 (Rel. 12, Last sequence update)				
DT 28-FEB-2003 (Rel. 41, Last annotation update)				

DNA-directed RNA polymerase beta" chain (BC 2.7.7.6).	QY	243	-----KVF-----PIEQFHVVGDPLITEPLPELOVERPLNPPRISQVNLYFN-	287
Spinacia olaracea (Spinach).	Db	754	YYQRITPTKKYEVLYRVVVPYETDGINLAT--LFPQDLQERDVNVLYNGC	811
Chloroplast.	QY	288	-ATRGF-----YCLRFNLLFNKCLAGS	308
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;	Db	812	KVTRGISDTSIQLVRLTCLVLN--WNQDKKGS	840
Spermatophyta; Magnoliophyta; eudicots; core eudicots;				
Caryophyllidae; Caryophyllales; Chenopodiaceae; Spinacia.				
NCBI TaxID=3562;				
[1]				
SEQUENCE FROM N.A.				
MEDLINE=8816331; PubMed=3045324;				
Hudson G.S.; Holton T.A.; Whitfield P.R.; Bottomley W.;				
Schmitz-Linnebecker C.; Maier R.M.; Alcaraz J.-P.; Cottet A.,				
Hermann R.G.; Mache R.;				
"The plastid chromosome of spinach (Spinacia oleracea) : complete nucleotide sequence and gene organization.";				
Plant Mol. Biol. 45:307-315 (2001).				
-1- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS SUBSTRATES.				
-1- CATALYTIC ACTIVITY: Nucleoside triphosphate = N diphosphate +				
-1- SUBUNIT: In chloroplasts, the RNA polymerase is composed of four subunits: alpha, beta, beta', and beta".				
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license (see <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to license@isb-sib.ch).				
EMBL; Au400848; CAB08715.1; -.				
PUB: A29559; A29559.				
HSSP; Q9XWU6; 1HQM.				
InterPro; IPR000722; RNA_Pol_A.				
InterPro; IPR007066; RNA_Pol_Rpb1_3.				
InterPro; IPR007083; RNA_Pol_Rpb1_4.				
InterPro; IPR007081; RNA_Pol_Rpb1_5.				
Pfam; PF04983; RNA_Pol_Rpb1_3; 1.				
Pfam; PF05000; RNA_Pol_Rpb1_4; 1.				
Pfam; PF04998; RNA_Pol_Rpb1_5; 1.				
Transferase; Transcriptase; DNA-directed RNA Polymerase; Chloroplast.				
SEQUENCE 1361 AA; 154768 MW; 19FF8C42BB1B17EB CRC64;				
Query Match 5.2%; Score 94.5%; DB 1; Length 1361;				
Best Local Similarity 20.8%; Pred No. 9.2;				
Matches 69; Conservative 52; Mismatches 101; Indels 109; Gaps 19;				
b 26 YLVARYGSSLRQLQPICPIEGRGGARTQAELLRLAQKFRGLHLF----RKGNASKE 79				
b 571 FIIIPPGSOPRKKEKMSLG----ISIETTNGFRKNSFAYDDPFRKRKSS---620				
b 80 QVRLLHDVQLPKAIIIGVKKGTRALLEMLNLHPAVVKASeIHFPNDENYGRGLEWY 139				
b 621 -----GITYKTG----IEMH-SIVKEKLIEY----RGYKEF 648				
b 140 RKKMPPSYPOOFTTIEKSPAYFTEEV----PERIYKMNSSIKLILIVREPTTAAISDX 193				
b 649 RPK----VOMKVDR----FFFPEEVHLAGSSIMVRNSLIGDWTINTRSIGGV 700				
b 194 TQVLEGKERKNTKYKPEKLAID----PTCEVNTKYKAV----RTSITYKHLERWL 242				
b 701 VRY----ERSKK----KIELTIESGDTGEGDKTSGTGHSGT----LPPSRKNSKDSKNNKWT 753				

Result No.	Score	Query	Match	Length	DB	ID	Description	SUMMARIES			
1	1814	Q96QIS homo sapien	Q8iz8	100.0	346	4	Q81ZT8	Query Match	100 %	Score 1814;	DB 4;
2	1768	Q96QIS homo sapien	Q8bz14	97.5	346	11	Q8BSI4	Best Local Similarity	100 %	Pred. No. 8.5-137;	Gaps 0;
3	935	Q96QIS homo sapien	Q8bz14	51.5	175	4	Q8BSI4	Matches	346;	Conservative 0;	Mismatches 0;
4	780	Q96QIS homo sapien	Q8bz14	43.0	307	4	Q14792	Qy	1	MLFKQQAWLQRQLVQSLVGSLLIVARYGSLDRLQPICPIEGRGGGARTQAEFPFLRA 60	61 LQFKQQAWLQRQLVQSLVGSLLIVARYGSLDRLQPICPIEGRGGGARTQAEFPFLRA 60
5	757.5	Q96QIS homo sapien	Q8bz14	41.8	311	11	Q35310	Db	1	MLFKQQAWLQRQLVQSLVGSLLIVARYGSLDRLQPICPIEGRGGGARTQAEFPFLRA 60	61 LQFKQQAWLQRQLVQSLVGSLLIVARYGSLDRLQPICPIEGRGGGARTQAEFPFLRA 60
6	726.5	Q96QIS homo sapien	Q8bz14	41.1	311	11	Q9BSGS	Qy	1	MLFKQQAWLQRQLVQSLVGSLLIVARYGSLDRLQPICPIEGRGGGARTQAEFPFLRA 60	61 LQFKQQAWLQRQLVQSLVGSLLIVARYGSLDRLQPICPIEGRGGGARTQAEFPFLRA 60
7	646.5	Q96QIS homo sapien	Q8bz14	40.0	605	5	QBMRE7	Db	1	MLFKQQAWLQRQLVQSLVGSLLIVARYGSLDRLQPICPIEGRGGGARTQAEFPFLRA 60	61 LQFKQQAWLQRQLVQSLVGSLLIVARYGSLDRLQPICPIEGRGGGARTQAEFPFLRA 60
8	643.5	Q96QIS homo sapien	Q8bz14	35.6	384	5	Q9VWJ7	Qy	1	MLFKQQAWLQRQLVQSLVGSLLIVARYGSLDRLQPICPIEGRGGGARTQAEFPFLRA 60	61 LQFKQQAWLQRQLVQSLVGSLLIVARYGSLDRLQPICPIEGRGGGARTQAEFPFLRA 60
9	643.5	Q96QIS homo sapien	Q8bz14	35.5	335	4	QNDNC2	Qy	1	MLFKQQAWLQRQLVQSLVGSLLIVARYGSLDRLQPICPIEGRGGGARTQAEFPFLRA 60	61 LQFKQQAWLQRQLVQSLVGSLLIVARYGSLDRLQPICPIEGRGGGARTQAEFPFLRA 60
10	637.5	Q96QIS homo sapien	Q8bz14	35.1	406	4	Q9Y663	Qy	1	MLFKQQAWLQRQLVQSLVGSLLIVARYGSLDRLQPICPIEGRGGGARTQAEFPFLRA 60	61 LQFKQQAWLQRQLVQSLVGSLLIVARYGSLDRLQPICPIEGRGGGARTQAEFPFLRA 60
11	636.5	Q96QIS homo sapien	Q8bz14	35.1	390	4	Q9Y662	Db	1	MLFKQQAWLQRQLVQSLVGSLLIVARYGSLDRLQPICPIEGRGGGARTQAEFPFLRA 60	61 LQFKQQAWLQRQLVQSLVGSLLIVARYGSLDRLQPICPIEGRGGGARTQAEFPFLRA 60
12	635.5	Q96QIS homo sapien	Q8bz14	35.0	367	4	Q9Y278	Qy	1	MLFKQQAWLQRQLVQSLVGSLLIVARYGSLDRLQPICPIEGRGGGARTQAEFPFLRA 60	61 LQFKQQAWLQRQLVQSLVGSLLIVARYGSLDRLQPICPIEGRGGGARTQAEFPFLRA 60
13	615.5	Q96QIS homo sapien	Q8bz14	33.9	393	11	Q8BKNG	Qy	1	MLFKQQAWLQRQLVQSLVGSLLIVARYGSLDRLQPICPIEGRGGGARTQAEFPFLRA 60	61 LQFKQQAWLQRQLVQSLVGSLLIVARYGSLDRLQPICPIEGRGGGARTQAEFPFLRA 60
14	614.5	Q96QIS homo sapien	Q8bz14	33.9	390	11	Q9QZS6	Qy	1	MLFKQQAWLQRQLVQSLVGSLLIVARYGSLDRLQPICPIEGRGGGARTQAEFPFLRA 60	61 LQFKQQAWLQRQLVQSLVGSLLIVARYGSLDRLQPICPIEGRGGGARTQAEFPFLRA 60
15	608	Q96QIS homo sapien	Q8bz14	33.5	311	4	Q96RX7	Qy	1	MLFKQQAWLQRQLVQSLVGSLLIVARYGSLDRLQPICPIEGRGGGARTQAEFPFLRA 60	61 LQFKQQAWLQRQLVQSLVGSLLIVARYGSLDRLQPICPIEGRGGGARTQAEFPFLRA 60
16	601.5	Q96QIS homo sapien	Q8bz14	33.2	250	4	Q9Y661	Qy	1	MLFKQQAWLQRQLVQSLVGSLLIVARYGSLDRLQPICPIEGRGGGARTQAEFPFLRA 60	61 LQFKQQAWLQRQLVQSLVGSLLIVARYGSLDRLQPICPIEGRGGGARTQAEFPFLRA 60



Query Match	Score 780; DB 4;	Length 307;	KW SIGNAL; Transferase.
Best Local Similarity	43.0%;	Pred. No. 2.7e-54;	POTENTIAL.
Matches	153; Conservative	64; Mismatches	HEPARAN SULFATE D-GLUCOSAMINYL 3-O-SULFOTRANSFERASE-1.
b	16 LGSLAVGSLLVIVARYGSSLRILOPICTEGRGGARTAQFLRAOFKRGHLHEFKGN 75	Indels 24; Gaps 6;	FT FT FT FT
b	1 MAALLIGAVL-LVAQ---PQLVPSRSP-----AELCQELLRKAGTLQD---- 39	SEQUENCE 311 AA; 35899 MW; CRC64;	CHAIN 21 311
b	76 ASKEQVRLHLDVQQLPKAIITIGVRKGTRALLEMLNHPAVVKAQSEIHFNDENYGKG 135	Query Match 41.8%; Score 757.5; DB 11; Length 311;	SEQUENCE 143; Conservative 54; Mismatches 77; Indels 19; Gaps 3;
b	40 DYRDGTAPNGSAQQQLPQTITIGVRKGTRALLEMLNHPDYAAANEVHFTWEEHYSHG 99	Best Local Similarity 48.8%; Pred. No. 1.7e-52;	QY 69 HEPRKGNAKSEQVRLHDIV-----
b	136 IEWYRKOMPFSYSPQQITTEKSPPAYFITEEVEPERIYKONSTKLILIVTREPTTRAISDYTQ 195	Matches 143; Conservative 54; Mismatches 77; Indels 19; Gaps 3;	21 HPAAPGPQLKQQLKQQLRKVILIPEDTGEETASNGSTOOLPQTITIGVRKGTRALLEMLNHL 80
b	100 LGWILSQMPFSYSPHQLVITKEPPAYFTSPKVPERVSYNSPFLRLDPSRVLSDYTQ 159	QY 113 HPAVYVRAQSQEIHFDNDENYGKIEWVKRMPPSYSPQQITIEKSPPAYFITEEVEPERIYKON 172	
b	196 VLEGKERKKNKTYKFEKLAIDPNTCEYNTKYKAVRTSYYTKHELMVLYKPELEQFTIVDQ 255	DB 81 HDPVAAEAEENEVEFFDWEEHYSQLGLGWLTQMPFSSPHQLTVKTPAYFTSPKVPERIYKSM 140	
b	160 VFYNNHOKHCKPKPSIERTFLVPGD--RANVDYCALNRSLYHVQNWNURFPARHHTIVDQ 217	QY 173 NSSIKILLIVREPTTRALSDYTQLEGERKERNKTYKFEKLAIDPNTCEYNTKYKAVRTS 232	
b	256 DRLLTPLPLQLOVERPLNLPPRISQVNLYFNATRGYCLRPNNIFPNMKCLASSKGRIHPE 315	DB 141 NPTIRLILRLRDPSERVLSDDYTQVNLNQLKHAPPFPEDLNRDG--RLNDYKALNES 198	
b	218 DRLLRPEPEIQKVERSLKLSPOINAQNFSYENKTGKBYCLR--DSGRDRCLLIESKGKAHQ 276	QY 233 IYTKHLERWIKYKPEIQLFHVYDQDRLLTPELPLQVLEKFLNUPRISQVNLYFNATRGYCL 292	
b	316 VDPSVITKLKEFHFPHNPKQFYQITGRTLNW 345	DB 199 LYTAHMNWLRFEPPLGHITHVODDLRDPFPIQKVERFLKSPQINAQNFSYENKTGKBYCLR 258	
b	277 VDPKLINKLKEYFHEPNPKFFELVGRFEDW 306	QY 293 YCRFNF1LNPKCLAGSKGRTHPVDPSVITKLKEFHFPHNPKQFYQITGRTLNW 345	
b	RESULTS 5	DB 259 YCR-DSGKDORCHESKGRAHPQVDPKCLDKLHEYFEPNPKCFEKFLVGETDW 310	
D	O35310 PRELIMINARY; PRT; 311 AA.	RESULTS 6	Q2ESG5 PRELIMINARY; PRT; 311 AA.
C	035310; PRELIMINARY; PRT; 311 AA.	C	Q9RSG5 PRELIMINARY; PRT; 311 AA.
C	01-JAN-1998 (T=EMBLrel. 05, Created)	C	Q9RSG5 PRELIMINARY; PRT; 311 AA.
T	01-JAN-1998 (T=EMBLrel. 05, Last sequence update)	AC	Q9RSG5 PRELIMINARY; PRT; 311 AA.
T	01-MAR-2003 (T=EMBLrel. 05, Last annotation update)	DT	01-MAR-2001 (T=EMBLrel. 16, Created)
T	Heparan sulfate D-glucosaminyl 3-O-sulfotransferase-1 precursor	DT	01-MAR-2001 (T=EMBLrel. 16, Last sequence update)
E	(Heparan sulfate (Glucosamine) 3-O-sulfotransferase 1).	DT	01-JUN-2002 (T=EMBLrel. 21, Last annotation update)
S	HS3STR OR 3OST1.	DB	3-O-sulfotransferase.
S	Mus musculus (Mouse).	OS	Rattus norvegicus (Rat).
C	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
C	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
X	NCBI_TaxID=10090; [1]	NCBI_TaxID=10116; [1]	NCBI_TaxID=10116; [1]
RN	SEQUENCE FROM N.A.	RN	SEQUENCE FROM N.A.
RP	SEQUENCE FROM N.A.	RP	SEQUENCE FROM N.A.
RC	STRAIN=C3H/An;	RC	STRAIN=Nistar;
RA	MEDLINE=98010647; PubMed=9346953;	RA	Li Z.-Y., Hirayoshi K., Suzuki Y.;
RA	Shwartz N.W., Liu J.J., Fritze L.M.S., Schwartz J.J., Zhang L.,	RA	"Expression of N-deacetylase/sulfotransferase and 3-O-sulfotransferase
RA	Loyerart D., Rosenberg R.D.;	RA	"in rat alveolar type II cells."
A	"Molecular cloning and expression of mouse and human cDNAs encoding	RL	An. J. Physiol. 279:L292-L301 (2000).
T	heparan sulfate D-glucosaminyl 3-O-sulfotransferase.";	DR	EMBL; AF177430; AAG09283.1; -.
T	J. Biol. Chem. 272:28019-28019 (1997).	DR	HSSP; P52848; INST.
L	[2] SEQUENCE FROM N.A.	DR	InterPro; IPR000863; Sulfotransferase.
N	SEQUENCE FROM N.A.	KW	Pfam; PF00655; Sulfotransferase.
P	SEQUENCE FROM N.A.	SEQUENCE 311 AA; 3581 MW; ACEE78B365BBCB75 CRC64;	SEQUENCE FROM N.A.
C	TISSUE=Breast tumor;	SEQUENCE 311 AA; 3581 MW; ACEE78B365BBCB75 CRC64;	SEQUENCE FROM N.A.
A	Strausberg R.; Submitted (JUN-2001) to the EMBL/GenBank/DDBJ databases.	SEQUENCE 311 AA; 3581 MW; ACEE78B365BBCB75 CRC64;	SEQUENCE FROM N.A.
N	[3] SEQUENCE FROM N.A.	SEQUENCE 311 AA; 3581 MW; ACEE78B365BBCB75 CRC64;	SEQUENCE FROM N.A.
P	STRAIN=C57BL/6J; TISSUE=Ovary;	SEQUENCE 311 AA; 3581 MW; ACEE78B365BBCB75 CRC64;	SEQUENCE FROM N.A.
X	MEDLINE=22354683; PubMed=12466881;	SEQUENCE 311 AA; 3581 MW; ACEE78B365BBCB75 CRC64;	SEQUENCE FROM N.A.
X	The FANTOM Consortium,	SEQUENCE 311 AA; 3581 MW; ACEE78B365BBCB75 CRC64;	SEQUENCE FROM N.A.
X	The RIKEN Genome Exploration Research Group Phase I & II Team;	SEQUENCE 311 AA; 3581 MW; ACEE78B365BBCB75 CRC64;	SEQUENCE FROM N.A.
X	"Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs.";	SEQUENCE 311 AA; 3581 MW; ACEE78B365BBCB75 CRC64;	SEQUENCE FROM N.A.
EMLB	EMBL; AF0152385; AAB84387.1; -.	QY 69 HEPRKGNAKSEQVRLHDIV-----	SEQUENCE 311 AA; 3581 MW; ACEE78B365BBCB75 CRC64;
EMLB	BC0091133; AAB09133.1; -.	21 HPAAPGPQLKQQLRKVILIPEDTGEETASNGSTOOLPQTITIGVRKGTRALLEMLNHL 80	SEQUENCE 311 AA; 3581 MW; ACEE78B365BBCB75 CRC64;
EMLB	P52848; INST.	113 HPAVYVRAQSQEIHFDNDENYGKIEWVKRMPPSYSPQQITIEKSPPAYFITEEVEPERIYKON 172	SEQUENCE 311 AA; 3581 MW; ACEE78B365BBCB75 CRC64;
HSSP	HSSP; MGII:1201606; HS-est1.	81 HDPVAAEAEENEVEFFDWEEHYSQLGLGWLTQMPFSSPHQLTVKTPAYFTSPKVPERIYKSM 140	SEQUENCE 311 AA; 3581 MW; ACEE78B365BBCB75 CRC64;
InterPro	InterPro; IPR000863; Sulfotransferase.	141 NPTIRLILRLRDPSERVLSDDYTQVNLNHLQKHKPPPTEDLNRDG--RLNDYKALNES 198	SEQUENCE 311 AA; 3581 MW; ACEE78B365BBCB75 CRC64;
Pfam	Pfam; PF000863; Sulfotransferase.	141 NPTIRLILRLRDPSERVLSDDYTQVNLNHLQKHKPPPTEDLNRDG--RLNDYKALNES 198	SEQUENCE 311 AA; 3581 MW; ACEE78B365BBCB75 CRC64;

RESULT 7						
Qy	Db	Qy	Db	Qy	Db	Qy
233 IYTCHKHLLERWLFYTFPIEQFHVDGDRLLITEPLPLOVEKPLNLPPIRQSONYLNFATRGF :		199 LYHARMLANLWRFFPLGEHTIVDGFREFDPRPEIYKVERFLCSPOINASNFTYKTKRGF :				
293 YCLRENLLENKCLAGSKGRTHPEVDPSTVKLRKEFHPNQKFQITGRTLNN :		259 YCLR - DSGKDRCLHESKGRAFPQDVDEKLXKLHEYFREPNNKKFPLVGRTEDW 345				
Qy	Qy	Qy	Qy	Qy	Qy	Qy
QBMRP7	QBMRP7	PRELIMINARY;	PRELIMINARY;	PRT;	PRT;	PRT;
ID	ID			605 AA.		
AC	AC					
DT	DT	(TREMBLrel.	22,			
01-OCT-2002	Created)					
DT	DT	(TREMBLrel.	22,			
01-MAR-2003	Last annotation update)					
DT	DT	(TREMBLrel.	23,			
01-MAR-2003	Last annotation update)					
DE	DE					
GH2068P.	CG15075.					
GN	OS	Drosophila melanogaster (Fruit fly).				
	OC	Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;				
	OC	Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;				
	OC	Ephydriidae; Drosophilidae; Drosophila.				
	CX	NCBI_TaxID:7227; RN				
	RN	[1]				
	RP	SEQUENCE FROM N.A.				
	RC	STRAIN=Berkely;				
	RA	Strapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,				
	RA	Champagne M., Chavez C., Dorsett V., Dresiek D., Farfan D., Frise E.,				
	RA	Gonzalez M., Guarini H., Kronmiller B., Li P., Liao G.,				
	RA	Miranda A., Mongal C.J., Munro J., Paragis V., Park S.,				
	RA	Patel S., Phouanenavong S., Wan X., Yu C., Lewis S.E., Rubin G.M.,				
	RA	Celniker S.				
	RL	Submitted (JUN-2002) to the EMBL/GenBank/DDBJ databases.				
	DR	EMBL; AY121626; AAH51953.1; -.				
	DR	FlyBase; FBgn0034385; CG15075.				
	SQ	SEQUENCE 605 AA; 66437 MW; 128E17C938DA19C2 CRC64;				
Query	Query	Query	Query	Query	Query	Query
Match	Match	Match	Match	Match	Match	Match
Best Local Similarity	Score 726.5;	Score 726.5;	Score 726.5;	Score 726.5;	Score 726.5;	Score 726.5;
Matches 152;	Pred. No. 1..2e-49;	Mismatches 62;	Indels 109;	Gaps 4		
Conservative						
Query	Query	Query	Query	Query	Query	Query
Match	Match	Match	Match	Match	Match	Match
Best Local Similarity	40.0 %;	Score 726.5;	DB 5;	Length 605;		
Matches 152;	Pred. No. 1..2e-49;	Mismatches 62;	Indels 109;	Gaps 4		
Conservative						
Query	Query	Query	Query	Query	Query	Query
Match	Match	Match	Match	Match	Match	Match
Best Local Similarity	41.5 %;	Score 726.5;	DB 5;	Length 605;		
Matches 152;	Pred. No. 1..2e-49;	Mismatches 62;	Indels 109;	Gaps 4		
Conservative						
Query	Query	Query	Query	Query	Query	Query
Match	Match	Match	Match	Match	Match	Match
Best Local Similarity	40.0 %;	Score 726.5;	DB 5;	Length 605;		
Matches 152;	Pred. No. 1..2e-49;	Mismatches 62;	Indels 109;	Gaps 4		
Conservative						
Query	Query	Query	Query	Query	Query	Query
Match	Match	Match	Match	Match	Match	Match
Best Local Similarity	41.5 %;	Score 726.5;	DB 5;	Length 605;		
Matches 152;	Pred. No. 1..2e-49;	Mismatches 62;	Indels 109;	Gaps 4		
Conservative						
Query	Query	Query	Query	Query	Query	Query
Match	Match	Match	Match	Match	Match	Match
Best Local Similarity	41.5 %;	Score 726.5;	DB 5;	Length 605;		
Matches 152;	Pred. No. 1..2e-49;	Mismatches 62;	Indels 109;	Gaps 4		
Conservative						
Query	Query	Query	Query	Query	Query	Query
Match	Match	Match	Match	Match	Match	Match
Best Local Similarity	40.0 %;	Score 726.5;	DB 5;	Length 605;		
Matches 152;	Pred. No. 1..2e-49;	Mismatches 62;	Indels 109;	Gaps 4		
Conservative						
Query	Query	Query	Query	Query	Query	Query
Match	Match	Match	Match	Match	Match	Match
Best Local Similarity	41.5 %;	Score 726.5;	DB 5;	Length 605;		
Matches 152;	Pred. No. 1..2e-49;	Mismatches 62;	Indels 109;	Gaps 4		
Conservative						
Query	Query	Query	Query	Query	Query	Query
Match	Match	Match	Match	Match	Match	Match
Best Local Similarity	40.0 %;	Score 726.5;	DB 5;	Length 605;		
Matches 152;	Pred. No. 1..2e-49;	Mismatches 62;	Indels 109;	Gaps 4		
Conservative						
Query	Query	Query	Query	Query	Query	Query
Match	Match	Match	Match	Match	Match	Match
Best Local Similarity	41.5 %;	Score 726.5;	DB 5;	Length 605;		
Matches 152;	Pred. No. 1..2e-49;	Mismatches 62;	Indels 109;	Gaps 4		
Conservative						
Query	Query	Query	Query	Query	Query	Query
Match	Match	Match	Match	Match	Match	Match
Best Local Similarity	40.0 %;	Score 726.5;	DB 5;	Length 605;		
Matches 152;	Pred. No. 1..2e-49;	Mismatches 62;	Indels 109;	Gaps 4		
Conservative						
Query	Query	Query	Query	Query	Query	Query
Match	Match	Match	Match	Match	Match	Match
Best Local Similarity	41.5 %;	Score 726.5;	DB 5;	Length 605;		
Matches 152;	Pred. No. 1..2e-49;	Mismatches 62;	Indels 109;	Gaps 4		
Conservative						
Query	Query	Query	Query	Query	Query	Query
Match	Match	Match	Match	Match	Match	Match
Best Local Similarity	40.0 %;	Score 726.5;	DB 5;	Length 605;		
Matches 152;	Pred. No. 1..2e-49;	Mismatches 62;	Indels 109;	Gaps 4		
Conservative						
Query	Query	Query	Query	Query	Query	Query
Match	Match	Match	Match	Match	Match	Match
Best Local Similarity	41.5 %;	Score 726.5;	DB 5;	Length 605;		
Matches 152;	Pred. No. 1..2e-49;	Mismatches 62;	Indels 109;	Gaps 4		
Conservative						
Query	Query	Query	Query	Query	Query	Query
Match	Match	Match	Match	Match	Match	Match
Best Local Similarity	40.0 %;	Score 726.5;	DB 5;	Length 605;		
Matches 152;	Pred. No. 1..2e-49;	Mismatches 62;	Indels 109;	Gaps 4		
Conservative						
Query	Query	Query	Query	Query	Query	Query
Match	Match	Match	Match	Match	Match	Match
Best Local Similarity	41.5 %;	Score 726.5;	DB 5;	Length 605;		
Matches 152;	Pred. No. 1..2e-49;	Mismatches 62;	Indels 109;	Gaps 4		
Conservative						
Query	Query	Query	Query	Query	Query	Query
Match	Match	Match	Match	Match	Match	Match
Best Local Similarity	40.0 %;	Score 726.5;	DB 5;	Length 605;		
Matches 152;	Pred. No. 1..2e-49;	Mismatches 62;	Indels 109;	Gaps 4		
Conservative						
Query	Query	Query	Query	Query	Query	Query
Match	Match	Match	Match	Match	Match	Match
Best Local Similarity	41.5 %;	Score 726.5;	DB 5;	Length 605;		
Matches 152;	Pred. No. 1..2e-49;	Mismatches 62;	Indels 109;	Gaps 4		
Conservative						
Query	Query	Query	Query	Query	Query	Query
Match	Match	Match	Match	Match	Match	Match
Best Local Similarity	40.0 %;	Score 726.5;	DB 5;	Length 605;		
Matches 152;	Pred. No. 1..2e-49;	Mismatches 62;	Indels 109;	Gaps 4		
Conservative						
Query	Query	Query	Query	Query	Query	Query
Match	Match	Match	Match	Match	Match	Match
Best Local Similarity	41.5 %;	Score 726.5;	DB 5;	Length 605;		
Matches 152;	Pred. No. 1..2e-49;	Mismatches 62;	Indels 109;	Gaps 4		
Conservative						
Query	Query	Query	Query	Query	Query	Query
Match	Match	Match	Match	Match	Match	Match
Best Local Similarity	40.0 %;	Score 726.5;	DB 5;	Length 605;		
Matches 152;	Pred. No. 1..2e-49;	Mismatches 62;	Indels 109;	Gaps 4		
Conservative						
Query	Query	Query	Query	Query	Query	Query
Match	Match	Match	Match	Match	Match	Match
Best Local Similarity	41.5 %;	Score 726.5;	DB 5;	Length 605;		
Matches 152;	Pred. No. 1..2e-49;	Mismatches 62;	Indels 109;	Gaps 4		
Conservative						
Query	Query	Query	Query	Query	Query	Query
Match	Match	Match	Match	Match	Match	Match
Best Local Similarity	40.0 %;	Score 726.5;	DB 5;	Length 605;		
Matches 152;	Pred. No. 1..2e-49;	Mismatches 62;	Indels 109;	Gaps 4		
Conservative						
Query	Query	Query	Query	Query	Query	Query
Match	Match	Match	Match	Match	Match	Match
Best Local Similarity	41.5 %;	Score 726.5;	DB 5;	Length 605;		
Matches 152;	Pred. No. 1..2e-49;	Mismatches 62;	Indels 109;	Gaps 4		
Conservative						
Query	Query	Query	Query	Query	Query	Query
Match	Match	Match	Match	Match	Match	Match
Best Local Similarity	40.0 %;	Score 726.5;	DB 5;	Length 605;		
Matches 152;	Pred. No. 1..2e-49;	Mismatches 62;	Indels 109;	Gaps 4		
Conservative						
Query	Query	Query	Query	Query	Query	Query
Match	Match	Match	Match	Match	Match	Match
Best Local Similarity	41.5 %;	Score 726.5;	DB 5;	Length 605;		
Matches 152;	Pred. No. 1..2e-49;	Mismatches 62;	Indels 109;	Gaps 4		
Conservative						
Query	Query	Query	Query	Query	Query	Query
Match	Match	Match	Match	Match	Match	Match
Best Local Similarity	40.0 %;	Score 726.5;	DB 5;	Length 605;		
Matches 152;	Pred. No. 1..2e-49;	Mismatches 62;	Indels 109;	Gaps 4		
Conservative						
Query	Query	Query	Query	Query	Query	Query
Match	Match	Match	Match	Match	Match	Match
Best Local Similarity	41.5 %;	Score 726.5;	DB 5;	Length 605;		
Matches 152;	Pred. No. 1..2e-49;	Mismatches 62;	Indels 109;	Gaps 4		
Conservative						
Query	Query	Query	Query	Query	Query	Query
Match	Match	Match	Match	Match	Match	Match
Best Local Similarity	40.0 %;	Score 726.5;	DB 5;	Length 605;		
Matches 152;	Pred. No. 1..2e-49;	Mismatches 62;	Indels 109;	Gaps 4		
Conservative						
Query	Query	Query	Query	Query	Query	Query
Match	Match	Match	Match	Match	Match	Match
Best Local Similarity	41.5 %;	Score 726.5;	DB 5;	Length 605;		
Matches 152;	Pred. No. 1..2e-49;	Mismatches 62;	Indels 109;	Gaps 4		
Conservative						
Query	Query	Query	Query	Query	Query	Query
Match	Match	Match	Match	Match	Match	Match
Best Local Similarity	40.0 %;	Score 726.5;	DB 5;	Length 605;		
Matches 152;	Pred. No. 1..2e-49;	Mismatches 62;	Indels 109;	Gaps 4		
Conservative						
Query	Query	Query	Query	Query	Query	Query
Match	Match	Match	Match	Match	Match	Match
Best Local Similarity	41.5 %;	Score 726.5;	DB 5;	Length 605;		
Matches 152;	Pred. No. 1..2e-49;	Mismatches 62;	Indels 109;	Gaps 4		
Conservative						
Query	Query	Query	Query	Query	Query	Query
Match	Match	Match	Match	Match	Match	Match
Best Local Similarity	40.0 %;	Score 726.5;	DB 5;	Length 605;		
Matches 152;	Pred. No. 1..2e-49;	Mismatches 62;	Indels 109;	Gaps 4		
Conservative						
Query	Query	Query	Query	Query	Query	Query
Match	Match	Match	Match	Match	Match	Match
Best Local Similarity	41.5 %;	Score 726.5;	DB 5;	Length 605;		
Matches 152;	Pred. No. 1..2e-49;	Mismatches 62;	Indels 109;	Gaps 4		
Conservative						
Query	Query	Query	Query	Query	Query	Query
Match	Match	Match	Match	Match	Match	Match
Best Local Similarity	40.0 %;	Score 726.5;	DB 5;	Length 605;		
Matches 152;	Pred. No. 1..2e-49;	Mismatches 62;	Indels 109;	Gaps 4		
Conservative						
Query	Query	Query	Query	Query	Query	Query
Match	Match	Match	Match	Match	Match	Match
Best Local Similarity	41.5 %;	Score 726				

RESULT 10  
KEY663

SQ	SEQUENCE	367 AA;	41501 MW;	F63EACDD4721607C CRC64;
RN	SEQUENCE FROM N.A.			
RP	TISSUE-Liver;			
RX	MEDLINE=99143187; PubMed=9988767;			
RA	Shworak N.W., Liu J., Petros L.M., Zhang L., Kobayashi M., Copeland N.G., Jenkins R.D.;			
PA	"Multiple isoforms of heparan sulfate D-glucosaminyl 3-O-sulfotransferase. Isolation, characterization, and expression of human cDNAs and identification of distinct genomic loci.";			
JR	J. Biol. Chem. 274:5170-5174 (1999).			
EMBL	AF105377;			
DR	HSSP; P5248; INST.			
DR	Genew; HGNC:5198; HS3ST3B1.			
DR	InterPro; IPR000663; Sulfotransfer; 1.			
PFAM	PF00885; Sulfotransfer; 1.			
KW	Transferase; 390 AA; 43324 MW; 5C54C3B9836BC614 CRC64;			
SQ	SEQUENCE	390 AA;	43324 MW;	5C54C3B9836BC614 CRC64;
Query Match	Score 35.5%; DB 4;	Score 635.5%; DB 4;	Score 35.5%; DB 4;	Length 367;
Best Local Similarity	46.0%; Pred. No. 1.2e-42;	Conservative	46.0%; Pred. No. 1.2e-42;	
Matches	120; Indels 50; Mismatches 82;			
RT	Multiple isoforms of heparan sulfate D-glucosaminyl 3-O-sulfotransferase. Isolation, characterization, and expression of human cDNAs and identification of distinct genomic loci.";			
RL	J. Biol. Chem. 274:5170-5174 (1999).			
DR	YQ	88 QQLPKAIIIGYRKGGTRALLEMINLHPAVVKASOEIHFDNDENYGKIEWRYKMPFSDYV	147	
DR	Db	112 KRLFQALIVGVRKGTRAVLEFRVHPDTRALGTEPHFS--DRYRGRLDWRSMPRTL	169	
DR	QY	148 PQQTIEKSPAYFTEVPERIYKMNNSKLILLIVREPTRAISDTYQVLEGKRKNTY	207	
DR	Db	170 ESQTILEKTPSYFTQEAQPRIFMMSRDTKLIVVVRNPTRALSDYTQTS---KKPDI	225	
DR	QY	208 YKEFKLAIKDNPNTCE-VNITYKVAFTSITYKHLRKYPIEOFHVVDGRLLTEPLPEL	266	
DR	Db	226 PTFRGLSFENRNLGLVLYDWSNAARIGMVTLHSLSWQYPLAQHFTVSSBRLLTDPAHEN	285	
DR	QY	267 QLYKEFKLNIIPRISQVNLYFNATRGFYCLCR--FNIENKCLAGSKGRHIPEVDPDSVITKL	324	
DR	Db	286 GRVQDFELGKRFITDKHRYFNKTKGFPCIKKTQSSLPLRCGSKGRTHVQIDBEVIDOL	345	
DR	QY	325 RKFHFPNQKPYQITGFLNW	345	
DR	Db	346 REFTYRPNKFYETVGQDFRW	366	
RESULT	13			
OB8KNG	1D OB8KNG6 PRELIMINARY; PRT; 393 AA.			
AC	OB8KNG6 PRELIMINARY; PRT; 393 AA.			
DT	01-MAR-2003 (TREMBLrel. 23, Created)			
DT	01-MAR-2003 (TREMBLrel. 23, Last sequence update)			
DT	01-MAR-2003 (TREMBLrel. 23, Last annotation update)			
DE	Similar to heparan sulfate D-Glucosaminyl 1 DE-3-O-sulfotransferase-3A.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Buteleostomi; Mammalia; Eutheria; Rodentia; Sciurognath; Muridae; Murinae; Mus.			
NCBI_TAXID	100900; NCBI_TAXID=100900;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=C57BL/6J; TISSUE=Dorsal root ganglion;			
RX	MEDLINE=22354683; PubMed=22466651;			
RA	The FANTOM Consortium,			
RA	the RIKEN Genome Exploration Research Group Phase I & II Team;			
RT	"Analysis of the mouse transcriptome based on functional annotation of full-length cDNAs";			
RT	Nature 420:563-573 (2002).			
RL	EMBL: AK051284; BAC34592.1; -			
DR	SEQUENCE 393 AA; 43483 MW; 0B3E6F6EBDB80412 CRC64;			
Query Match	Score 33.9%; DB 11; Length 393;			
Best Local Similarity	37.9%; Pred. No. 5.1e-41;			
Matches	142; Indels 60; Mismatches 128;			
RT	SEQUENCE FROM N.A.			
RC	09Y278 PRELIMINARY; PRT; 367 AA.			
AC	09Y278 PRELIMINARY; PRT; 367 AA.			
DT	01-NOV-1999 (TREMBLrel. 12, Created)			
DT	01-NOV-1999 (TREMBLrel. 12, Last sequence update)			
DT	01-OCT-2001 (TREMBLrel. 22, Last annotation update)			
DE	Heparan sulfate D-glucosaminyl 3-O-sulfotransferase-2 (EC 2.8.2.23).			
GN	3OST2.			
OS	Homo sapiens (Human). Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
NCBI_TAXID	9606; NCBI_TAXID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	09Y278 PRELIMINARY; PRT; 367 AA.			
RA	Shworak N.W., Liu J., Petros L.M., Zhang L., Kobayashi M., Copeland N.G., Jenkins R.D.;			
RA	"Multiple isoforms of heparan sulfate D-glucosaminyl 3-O-sulfotransferase. Isolation, characterization, and expression of human cDNAs and identification of distinct genomic loci.";			
RT	J. Biol. Chem. 274:5170-5184 (1999).			
DR	YQ	10 RQKLIVGSLAVGSLLVARYVSLDLOQPIPIEGR--LGARTQFPLALQFGRGL	67	
DR	Db	23 RKFLMLCSLTSLYVYCYLAERCPSGSPYAGVPGFGRVPGPRELMWNPAGPKRLQJ	82	
DR	QY	68 LHEFRK----GNASEQ-----VRLHDLVQOLPKA	93	
DR	Db	83 LRQRRRGRGSGPQGDSSQQEEQSPGLANAPGSGAGSSVAEAOPTLMLLDGSKQDQQA	142	
DR	QY	94 IIGVKGTRGTRALLEMINLHPAVVKASOEIHFDNDENYGKIEWRYKMPFSDYV	153	
DR	Db	143 IIIIGVKGTRALLEFRVHPDTRALGTEPHFS--DRYRGRLDWRSMPRTL	200	
DR	QY	154 EKSPAYFITEEPERIYKMNNSKLILLIVREPTRAISDTYQVLEGKRKNTYKPEKL	213	
DR	Db	201 EKTPSVFVTREPARISAMSQDTKLIVVVRDPVTRASDYTQTS---KRPDIPSESL	256	
DR	KW	Transferase; 367 AA; 43483 MW; 0B3E6F6EBDB80412 CRC64;		
DR	QY	214 AIDPNTCE-VNITYKVAFTSITYKHLRKYPIEOFHVVDGRLLTEPLPEL-TTEPLPELQVKEF	273	

**RESULT 14**

Q9QZS6 PRELIMINARY; PRT; 390 AA.

ID Q9QZS6; PRELIMINARY;

AC Q9QZS6; Created)

DT 01-MAY-2000 (TREMBUREL. 13, Last sequence update)

DT 01-MAY-2000 (TREMBUREL. 13, Last annotation update)

DT 01-JUN-2002 (TREMBUREL. 21, Last annotation update)

DE HSSST3B.

OS Mus musculus (Mouse).  
OC Metazoa: Chordata; Craniata: Vertebrata; Euteleostomi;  
OC Mammalia: Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;

SEQUENCER FROM N.A.

STRAIN=C57BL/6J;

RX PubMed=10520990;

RA Shukla D., Liu J., Blaiklock P., Shwartz N.W., Bai X., Esko J.D., Cohen G.H., Eisenberg R.J., Rosenberg R.D., Spear P.G.; RT A novel role for 3-O-sulfated heparan sulfate in herpes simplex virus 1 entry";

RL Cell 99:13-22(1999).

DR AF04505.1; -.

DR HSSP; P58448; INST\_MGD; MGI\_33353; Ms3sc3b.

DR InterPro IPR000867; Sulfotransferase.

DR Pfam: PF00685; Sulfotransfer; 1.

KW Transferase.

SEQUENCE 390 AA; 43326 MW; ACD0D28D6B3DDE8 CRC64;

Query Match 33.9%; Score 614.5; DB 11; Length 390; Matches 127; Conservative 43; Nismatches 92; Indels 19; Gaps 5;

Best Local Similarity 45.2%; Pred. No. 6.1e-41; Job time : 44 secs

DB 125 ISSPFSGAGSK-----QLOAIIIVGKKKGTRALLEFLRVHDYAVGAEPHF- 173

QY 68 LHEFKGNASKEQYRVLHDLVQOLPKAATTCVKGCGTRALLEMLNHPAVKASOEHFFD 127

DB 128 NDENYGKGLIWFYRKMPFSPSYQQPIQTIKSPAYFTEEVPERIYKMSSSKKLLIVREPT 187

QY 129 :-----DRSHKGHLAWYRDLMPTKLGQT-----KRDIPSFSEULFRNSASLIDTWASAIIGLYAKHLPWLRIEP 232

DB 174 RAISDYTOLECKERKNKTYKPEKLAIDENTCE-VNTKKAVENTSITTKHLBWKYPF 245

DB 233 RAISPYTQTL-----KRDIPSFSEULFRNSASLIDTWASAIIGLYAKHLPWLRIEP 288

QY 188 1EOPHIVDGRLLITEPLPFLQYVFLNPIMEKPSYFTREAPARISAMSOKTKLIVVRDPT 246

DB 247 247 1EOPHIVDGRLLITEPLPFLQYVFLNPIMEKPSYFTREAPARISAMSOKTKLIVVRDPT 304

DB 289 LGQMLFVFSGERLVSDPAGEERRVQDFLGKURITTDKFYFNQIKGFPCLUKAEGSGKPHC 348

QY 305 LAGSKGRRIPEVDPSVITKARKKFFPENQKFYQITGRTLNW 345

DB 349 LGKTKGRAHPTIAREVLQRQDFTRPFNRKRFYQMTGRDFCW 389

**RESULT 15**

Q96XX7 PRELIMINARY; PRT; 311 AA.

ID Q96XX7; Created)

DT 01-DEC-2001 (TREMBUREL. 19, Last sequence update)

DT 01-DEC-2001 (TREMBUREL. 19, Last sequence update)

DB 01-OCT-2002 (TREMBUREL. 22, Last annotation update)

OS Heparan sulphate D-glucosaminyl 3-O-sulfotransfase-3B like.

OC Homo sapiens (Human).

OC Buiaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=2109310; PubMed=1115779;

RA Daniels R.J., Pedan J.F., Lloyd C., Horsley S.W., Clark K., Tuffarelli C., Kearney L., Buckle V.J., Doggett N.A., Flint J., RA Higgs D.R.; RT "Sequence, structure and pathology of the fully annotated terminal 2 Mb of the short arm of human chromosome 16.";

RT RL Human Mol. Genet. 10:339-352 (2001).

DR EMBL: AE006640; AAC61299.1; -.

DR InterPro: IPR000863; Sulfotransferase.

DR Pfam: PF00685; Sulfotransfer; 1.

KW Transferase.

SQ SEQUENCE 311 AA; 34694 MW; 6841E6151BA0DA6F CRC64;

Query Match 33.5%; Score 605; DB 4; Length 311;

Best Local Similarity 44.7%; Pred. No. 1.5e-40;

Matches 119; Conservative 46; Nismatches 81; Indels 20; Gaps 4;

QY 88 QQIPKAAIIGVYRGCGTRALLEMLNHPAVKASOEHFFDNDENYKKG3IEMWYRKKMPFSY 147

DB 57 RRFFQALIVGvKKGTRALLEFLRHPDYLSEPHF--DRCYERGIAWTRSLMPRTL 114

QY 148 PQQTTIEPSPAYFTEEPERIYKMSSSKKLLIVREPTRASTDYQTELEGKRKNKY 207

DB 115 DGQTMTMEKTPSYFTVRAAPRTHAMSPDTKLIVVVRNVTRALSDYTQTLS---KTPGJ 170

QY 208 YKFKEKLADPNTCEVNTKYKAVTSITTKHLBWKYPFQEHHVGDRLITEPLPFLQ 267

DB 171 PSFRALAFTRHGLGPVDTAWSAVIGLWAZCHLDWLRHPLSHFLVFSGERLVSDPAGEVGC 230

QY 268 LVEKFLNLPPRISOYNNYFNAATGFYCILRNLFNK-----CLAGSKGRTHPEVDPF 319

DB 231 RVQDFGLKRRVVTDKHFFENATKGFPCLK-----KAOGGSRPRCLGKSKGRDHPRVQAA 284

QY 320 VTKLRFKFFPENQKFYQITGRTLNW 345

DB 285 VVERLQEFTRPENRFRYQMTGQDEGW 310

Search completed: December 4, 2003, 16:44:00

Gencore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw mode.

Run on: December 4, 2003, 16:43:06 ; Search time 29 Seconds

(without alignments)  
504,812 Million cell updates/sec

Title: PCT-US03-21094-2

Perfect score: 1814

Sequence: 1 MLFKQOAWLRQKLLVIGSIA.....FFHPFNQKFYQITGRTLNWMP 346

Scoring table: BLOSUM62

Gapext: 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:\*

1: /cgn2\_6/\_ptodata/1/iaa/5A\_COMBO.PEP:\*

2: /cgn2\_6/\_ptodata/1/iaa/5B\_COMBO.PEP:\*

3: /cgn2\_6/\_ptodata/1/iaa/6A\_COMBO.PEP:\*

4: /cgn2\_6/\_ptodata/1/iaa/6B\_COMBO.PEP:\*

5: /cgn2\_6/\_ptodata/1/iaa/PCTNS\_COMBO.PEP:\*

6: /cgn2\_6/\_ptodata/1/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	164.6	90.7	321	4 US-09-735-935-4	Sequence 2, App1
2	772.5	42.6	255	4 US-09-735-935-4	Sequence 4, App1
3	9.9	5.5	803	1 US-08-062-368-4	Sequence 4, App1
4	95.5	5.3	1088	2 US-08-742-026-2	Sequence 2, App1
5	95.5	5.3	1088	2 US-08-742-026-23	Sequence 23, App1
6	9.1	5.0	1525	3 US-09-396-651B-1	Sequence 1, App1
7	90.5	5.0	1076	4 US-09-134-001C-4037	Sequence 4037, App1
8	88.5	4.9	473	4 US-09-134-001C-4200	Sequence 4200, App1
9	88.5	4.9	600	4 US-09-996-243-347	Sequence 347, App1
10	88	4.9	383	4 US-09-10-532A-5773	Sequence 5773, App1
11	87.5	4.8	296	4 US-09-134-001C-2976	Sequence 2976, App1
12	87	4.8	431	4 US-09-107-532A-5796	Sequence 5796, App1
13	86.5	4.8	283	2 US-08-332-562A-4200	Sequence 136, App1
14	86.5	4.8	426	4 US-09-134-001C-3346	Sequence 3346, App1
15	86.5	4.8	467	4 US-09-134-001C-4200	Sequence 4200, App1
16	86.5	4.8	3898	3 US-08-750-717-2	Sequence 2, App1
17	86	4.7	411	4 US-09-328-352-6424	Sequence 6424, App1
18	86	4.7	757	4 US-09-107-532A-7189	Sequence 7189, App1
19	86	4.7	1211	4 US-09-328-352-7967	Sequence 7967, App1
20	85.5	4.7	395	4 US-09-252-991A-24415	Sequence 24415, App1
21	85.5	4.7	426	3 US-08-676-444-40	Sequence 40, App1
22	85	4.7	555	4 US-09-461-325-251	Sequence 251, App1
23	85	4.7	1307	4 US-09-252-991A-20867	Sequence 20867, App1
24	84	4.6	345	4 US-09-328-352-7063	Sequence 7063, App1
25	83.5	4.6	222	4 US-09-255-991A-26656	Sequence 26656, App1
26	83	4.6	455	4 US-09-107-532A-6755	Sequence 6755, App1
27	83	4.6	460	4 US-09-134-001C-2979	Sequence 2979, App1

#### ALIGNMENTS

RESULT 1  
US-09-735-935-2  
; Sequence 2, Application US/09735935  
; Patent No. 6420150  
; GENERAL INFORMATION:  
; APPLICANT: GUEGLIER, Karl et al.  
; TITLE OF INVENTION: ISOLATED HUMAN DRUG-METABOLIZING PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN DRUG-METABOLIZING PROTEINS,  
; TITLE OF INVENTION:  
; TITLE OF INVENTION:  
; FILE REFERENCE: CL000970  
; CURRENT APPLICATION NUMBER: US/09-735,935  
; CURRENT FILING DATE: 2000-12-14  
; NUMBER OF SEQ ID NOS: 4  
; SEQ ID NO: 2  
; LENGTH: 321  
; TYPE: PRT  
; ORGANISM: Human  
US-09-735-935-2

Query Match 90.7%; Score 1646, DB 4; Length 321;  
Best Local Similarity 100.0%; Pred. No. 2.9e-164;  
Matches 311; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 36 RIQPICPIEGRIGGARTQAEPFLRAIQFKRLHLBPRGNASKEQRVLHDLVQQLPRAII 95  
Db 11 RIQPICPIEGRIGGARTQAEPFLRAQFKRLHLBPRGNASKEQRVLHDLVQQLPRAII 70  
Qy 96 IGVRKGCGTRALLEMLNHPAVTKASOBTHFDNDENYGGKIEWYRCKMPFSYPOQQTIEK 155  
Db 71 IGVRKGCGTRALLEMLNHPAVTKASOBTHFDNDENYGGKIEWYRCKMPFSYPOQQTIEK 130  
Qy 156 SPAFVTFEEVPERIYKONSSIKLILVPEPTBAISDYTQLEGKRNKTYYKPKYLAI 215  
Db 131 SPAFVTFEEVPERIYKONSSIKLILVPEPTRAISDYTQLEGKRNKTYYKPKYLAI 190  
Qy 216 DPNTCEVNTKYKAVRPSIYTKHRLWLYKPPJEQFFIVDGDRLITPLPEIQLVERFLNL 275  
Db 191 DPNTCEVNTKYKAVRPSIYTKHRLWLYKPPJEQFFIVDGDRLITPLPEIQLVERFLNL 250  
Qy 276 PPRISQNYLNATRGYCLRFNTIIFNKCLASGSKGRHPEVDPSVTKLRLKFFHPPNQKF 335  
Db 251 PPRISQNYLNATRGYCLRFNTIIFNKCLASGSKGRHPEVDPSVTKLRLKFFHPPNQKF 310  
Qy 336 YQITGTLNWNP 346  
Db 311 YQITGTLNWNP 321



CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Brook, David E.  
 REGISTRATION NUMBER: 22,592  
 REFERENCE/DOCKET NUMBER: CPI95-01  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 617-861-6240  
 TELEFAX: 617-861-9540  
 INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1088 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-742-026-2

Query Match Score 95.5; DB 2; Length 1088;  
 Best Local Similarity 19.8%; Pred. No. 0.73; Gaps 11;  
 Matches 44; Conservative 48; Mismatches 63; Indels 67; Gaps 11;

Qy 123 IHPFDNDENYGKGLEWY---RKKMPPSYPQQTIIEKS PAYFITEEVPERIYKNSSIKL 178  
 Db 731 LHFDLDTN-----WYTRFRNRR-----KGYASDDDEDTQKGLNTLVEA 770

Qy 123 THFFDNDENYGKGLEWY---RKKMPPSYPQQTIIEKS PAYFITEEVPERIYKNSSIKL 178  
 Db 771 LHLTL-----SRANAPFTPLADGYTORYKVFQEDLKEIAINPKNVLSVHFLSYPSVR 826

Qy 129 LITVREPTRAISDYTQVL-EGKERKNTYYK---FEKLAIDPNTCEVNT---KYKAVR 230  
 Db 771 LHLTL-----SRANAPFTPLADGYTORYKVFQEDLKEIAINPKNVLSVHFLSYPSVR 826

Qy 231 TSLTYKTHLE---RMLKXFPPIEOFHVYDGRILTE----PLPELQVEKFLNL--- 275  
 Db 827 QELFDKEVAYARMQK-----VDMARNTREKOMISLKTPNLEVLVLSADADLKD 878

Qy 276 -----PRRISQYNLYFNATRGFYCLRFNTIFNKCLAGSK 309  
 Db 879 IDS LKGYISDELNRVNNTSDEAKYCVYESCVADWPVLGKK 920

RESULT 6  
 US-09-396-651B-1  
 Sequence 1, Application US/09396651B  
 Patent No. 6223076  
 GENERAL INFORMATION:  
 APPLICANT: Darst, Seth A  
 APPLICANT: Zhang, Gongyi  
 APPLICANT: Campbell, Elizabeth  
 APPLICANT: Minakin, Leonid  
 APPLICANT: Severinov, Konstantin  
 TITLE OF INVENTION: A CRYSTAL OF BACTERIAL CORE RNA POLYMERASE AND METHODS  
 FILE REFERENCE: 600-1-58  
 CURRENT APPLICATION NUMBER: US/09/396,651B  
 CURRENT FILING DATE: 1999-09-15  
 NUMBER OF SEQ ID NOS: 4  
 SEQ ID NO 1  
 SOFTWARE: PatentIn Ver. 2.0  
 LENGTH: 1525  
 TYPE: PRT  
 ORGANISM: Thermus aquaticus  
 FEATURE:  
 NAME/KEY: SITE  
 LOCATION: (1247)  
 OTHER INFORMATION: Any amino acid can be at this position  
 US-09-396-651B-1

Query Match Score 91; DB 3; Length 1525;  
 Best Local Similarity 20.6%; Pred. No. 3.6; Gaps 8;  
 Matches 41; Conservative 31; Mismatches 65; Indels 62; Gaps 8;

Qy 85 DLVQQLPRAI-TIGVRKGCTRALLEMNLHPAVVKASOBHFDDNDENYGKGTIEWRKKM 143  
 Db 1252 DITQGLPRLVIELPEARFKAKAVISEDGUVRLLEGDRSLSVFSECGFSKE----- 1303

Qy 144 PPSYPQQTIIEKS PAYFITEEVPERIYKNSSIKL-----LIVREPTTRALSDYTVQL 197  
 Db 1304 -----YKLKDARLKVKGDYVEAGQPLTRGAIDPHQH 1337

Qy 198 EGKERKNTYYKKEKLAIDPNTCEVNTYKAVTSITYKLE---RWLKYPFIEQFHVV 253  
 Db 1338 EAKGPE----AVERYLVD---EIQKVRAGQVYKLIDKHIEIVVRQMLKVYEVTD--- 1384

US-08-742-026-23  
 Sequence 23, Application US/08/742026  
 Patent No. 5885815  
 GENERAL INFORMATION:  
 APPLICANT: Sassanfar, Mandana  
 APPLICANT: Kaufmann, Christoph  
 APPLICANT: Gallant, Paul L.  
 APPLICANT: Kranz, Janice E.  
 APPLICANT: Houman, Fariba  
 TITLE OF INVENTION: Candida Isoleucyl-tRNA Synthetase  
 TITLE OF INVENTION: Proteins, Nucleic Acids and Strains  
 NUMBER OF SEQUENCES: 23  
 CORRESPONDENCE ADDRESS:  
 STREET: Two Milliton Drive  
 CITY: Lexington  
 STATE: Massachusetts  
 COUNTY: USA  
 ZIP: 02173  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/742,026  
 FILING DATE:  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Brook, David E.  
 REGISTRATION NUMBER: 22,592  
 REFERENCE/DOCKET NUMBER: CPI95-01  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 617-861-6240

Qy 254 DGDRLITEPLPEROLVKEF 272  
 Db 1385 PGB----SPILLEQVLERW 1399

RESULT 7  
 US-09-134-001C-4037  
 / Sequence 4037, Application US/09134001C  
 / GENERAL INFORMATION:  
 / APPLICANT: Lynn Doucette-Stamm et al  
 / TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
 / FILE REFERENCE: GTC-007  
 CURRENT APPLICATION NUMBER: US/09/134,001C  
 PRIOR APPLICATION NUMBER: US 6,0/064,964  
 PRIOR FILING DATE: 1997-11-08  
 PRIOR APPLICATION NUMBER: US 6,0/055,779  
 PRIOR FILING DATE: 1997-08-14  
 SEQ ID NO 4 037  
 LENGTH: 1,076  
 TYPE: PRT  
 ORGANISM: *Staphylococcus epidermidis*  
 US-09-134-001C-4037

Query Match Similarity 5.0%; Score 90.5; DB 4; Length 1076;  
 Best Local Similarity 21.7%; Pred. No. 2,4;  
 Matches 58; Conservative 40; Mismatches 86; Indels 83; Gaps 14;

Qy 45 GRIGGARTOAE-FPL--PALQPK----RGLLHFP-----RKGNASKEQYQLRHLIVQQ 89  
 Db 723 GLIGGKGRANLYGLFKNAVEPENNSRFRGLQFIRFLDELDTRKCDKFGENV----- 774

Qy 90 LPKAIIITGVRKGTTRALLEMINH-----PAVKRASQETHFFDNDN-----YGR 134  
 Db 775 -----VGPNDNTVRMMTHSSKGLEFPVPIVGLSKKFNKGDLNAPVILNOQYGL 824

Qy 135 GIEWYRKMPFSYPOQQTTEKSPPAY-----FITEEVPERITYKNSKILLIVREPT 187  
 Db 825 GMDFDVNRDMAPPSSLAV-----AYRAINERELISEE-----MRLIYVA--LR 866

Qy 188 RAISDTQVLEGERKRNKTYKKFPEKLAIDPNCTCEVNTKYKAVRT----SITYTKHLERWL 242  
 Db 867 RAKEQL--ILVERVKDKSOLIXKEKLAVSDTIANNERLTATNPFLVIGTLAKHQSPSL 924

Qy 243 KYFPIQEFHVVDGDRLLTEPLPELQLV 269  
 Db 925 ---PNDQRFERDIDOLNSEKVERSVIV 948

CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/107,532A  
 FILING DATE: 30-Jun-1998

PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 60/085,598  
 FILING DATE: 14 May 1998  
 APPLICATION NUMBER: 60/051571  
 FILING DATE: July 2, 1997

ATTORNEY/AGENT INFORMATION:  
 NAME: Arinello, Pamela Denene

REGISTRATION NUMBER: 40,189  
 REFERENCE/DOCKET NUMBER: GTC-012  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (781)893-5007  
 TELEFAX: (781)893-8277

INFORMATION FOR SEQ ID NO: 4200:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 473 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 HYPOOTHECTICAL: YES  
 ORIGINAL SOURCE:  
 FEATURE:  
 NAME/KEY: misc\_feature  
 LOCATION: (B) LOCATION 1...473  
 SEQUENCE DESCRIPTION: SEQ ID NO: 4200:

US-09-107-532A-4200

Query Match Similarity 4.9%; Score 88.5; DB 4; Length 473;  
 Best Local Similarity 23.5%; Pred. No. 1,1;  
 Matches 50; Conservative 36; Mismatches 80; Indels 47; Gaps 10;

Qy 108 EMLNLIPAVVKAQSEIHF---FDNDENYGKIGIEWYRKKMPFSYPOQQTTEKS----PAY 159  
 Db 209 EKINIVQOLVTKKEKKFLNTTFFENLTMINNI-----MNFEDESIBASGFRQPEY 260

Qy 160 FITEEVPERITYKNSKILLIVREPT-----IVRPTPTTRALSDYTCVLEGERKRNKTYK 209  
 Db 261 ---ELFKRFLDITVSLILLLASPMFLTIALYKITSPGDIIYQVRLYKNQKEPSIYK 316

Qy 210 FEKLA--IDPNTCEVNTKYKAVRTSYKHLRMLKYLKFPLFQ-FHVVDGDRLLTEPLPEL 266

Db 317 FRMSATAEAKSGPVIAKSNDARVTPVGKET-RAYRFDELQPIFNVLGHDMSIVSERPE- 374

Qy 267 QLVERKLNLPPRISONLYNATRGFYCLRFNI 299  
 Db 375 -----RPFVFDQ---FNEENYYVYLHRHV 395

RESULT 9  
 US-09-996-243-347  
 / Sequence 347, Application US/0996243  
 / Patent No. 6478825  
 / GENERAL INFORMATION:  
 / APPLICANT: Ashkenazi, Avi J.  
 / APPLICANT: Baker, Kevin P.  
 / APPLICANT: Botstein, David  
 / APPLICANT: Desnoyers, Luc  
 / APPLICANT: Eaton, Dan L.  
 / APPLICANT: Ferrara, Napoléone  
 / APPLICANT: Fong, Sherman  
 / APPLICANT: Gerber, Hans Peter  
 / APPLICANT: Gerritsen, Mary E.  
 / APPLICANT: Goddard, Audrey  
 / APPLICANT: Godowski, Paul J.  
 / APPLICANT: Grimaldi, J. Christopher  
 / APPLICANT: Gurney, Austin L.  
 / APPLICANT: Klaivin, Ivar J.  
 / APPLICANT: Napiér, Mary A.  
 / APPLICANT: Pan, James  
 / APPLICANT: Paoni, Nicholas F.

RESULT 8  
 US-09-107-532A-4200  
 / Sequence 4200, Application US/09107532A  
 / Patent No. 6533275  
 / GENERAL INFORMATION:  
 / APPLICANT: Lynn A. Doucette-Stamm and David Bush  
 / TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTRICOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS  
 / NUMBER OF SEQUENCES: 7310  
 / CORRESPONDENCE ADDRESS:  
 ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
 STREET: 100 Beaver Street  
 CITY: Waltham  
 STATE: Massachusetts  
 COUNTRY: USA  
 ZIP: 02435  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: CD-ROM ISO9660  
 COMPUTER: PC  
 OPERATING SYSTEM: <Unknown>  
 SOFTWARE: ASCII

APPLICANT: Roy, Margaret Ann  
 APPLICANT: Stewart, Timothy A.  
 APPLICANT: Tumas, Daniel  
 APPLICANT: Watanabe, Colin K.  
 APPLICANT: Williams, P. Mickey L.  
 APPLICANT: Wood, William I.  
 APPLICANT: Zhang, Zemin  
 TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic acids Encoding the Same  
 FILE REFERENCE: P2730P1C13  
 CURRENT APPLICATION NUMBER: US/09/996,243  
 PRIOR APPLICATION NUMBER: 60/049787  
 PRIOR FILING DATE: 1997-06-16  
 PRIOR APPLICATION NUMBER: 60/062450  
 PRIOR FILING DATE: 1997-10-17  
 PRIOR APPLICATION NUMBER: 60/065186  
 PRIOR FILING DATE: 1997-11-12  
 PRIOR APPLICATION NUMBER: 60/065311  
 PRIOR FILING DATE: 1997-11-13  
 PRIOR APPLICATION NUMBER: 60/066770  
 PRIOR FILING DATE: 1997-11-24  
 PRIOR APPLICATION NUMBER: 60/075945  
 PRIOR FILING DATE: 1998-02-25  
 PRIOR APPLICATION NUMBER: 60/076910  
 PRIOR FILING DATE: 1998-03-20  
 PRIOR APPLICATION NUMBER: 60/083322  
 PRIOR FILING DATE: 1998-04-28  
 PRIOR APPLICATION NUMBER: 60/084600  
 PRIOR FILING DATE: 1998-05-07  
 PRIOR APPLICATION NUMBER: 60/087106  
 PRIOR FILING DATE: 1998-05-28  
 PRIOR APPLICATION NUMBER: 60/087607  
 PRIOR FILING DATE: 1998-06-02  
 PRIOR APPLICATION NUMBER: 60/087609  
 PRIOR FILING DATE: 1998-06-02  
 PRIOR APPLICATION NUMBER: 60/087759  
 PRIOR FILING DATE: 1998-06-02  
 PRIOR APPLICATION NUMBER: 60/087827  
 PRIOR FILING DATE: 1998-06-03  
 PRIOR APPLICATION NUMBER: 60/088021  
 PRIOR FILING DATE: 1998-06-04  
 PRIOR APPLICATION NUMBER: 60/088025  
 PRIOR FILING DATE: 1998-06-04  
 PRIOR APPLICATION NUMBER: 60/088026  
 PRIOR FILING DATE: 1998-06-04  
 PRIOR APPLICATION NUMBER: 60/088028  
 PRIOR FILING DATE: 1998-06-04  
 PRIOR APPLICATION NUMBER: 60/088029  
 PRIOR FILING DATE: 1998-06-04  
 PRIOR APPLICATION NUMBER: 60/088030  
 PRIOR FILING DATE: 1998-06-04  
 PRIOR APPLICATION NUMBER: 60/088033  
 PRIOR FILING DATE: 1998-06-04  
 PRIOR APPLICATION NUMBER: 60/088326  
 PRIOR FILING DATE: 1998-06-04  
 PRIOR APPLICATION NUMBER: 60/088167  
 PRIOR FILING DATE: 1998-06-05  
 PRIOR APPLICATION NUMBER: 60/088202  
 PRIOR FILING DATE: 1998-06-05  
 PRIOR APPLICATION NUMBER: 60/088312  
 PRIOR FILING DATE: 1998-06-05  
 PRIOR APPLICATION NUMBER: 60/088217  
 PRIOR FILING DATE: 1998-06-05  
 PRIOR APPLICATION NUMBER: 60/088655  
 PRIOR FILING DATE: 1998-06-09  
 PRIOR APPLICATION NUMBER: 60/088734  
 PRIOR FILING DATE: 1998-06-10  
 PRIOR APPLICATION NUMBER: 60/088738  
 PRIOR FILING DATE: 1998-06-10  
 PRIOR APPLICATION NUMBER: 60/088742  
 PRIOR FILING DATE: 1998-06-10  
 PRIOR APPLICATION NUMBER: 60/088810  
 ; PRIOR FILING DATE: 1998-06-11  
 ; PRIOR APPLICATION NUMBER: 60/088824  
 ; PRIOR FILING DATE: 1998-06-10  
 ; PRIOR APPLICATION NUMBER: 60/088826  
 ; PRIOR FILING DATE: 1998-06-10  
 ; PRIOR APPLICATION NUMBER: 60/088858  
 ; PRIOR FILING DATE: 1998-06-11  
 ; PRIOR APPLICATION NUMBER: 60/088861  
 ; PRIOR FILING DATE: 1998-06-11  
 ; PRIOR APPLICATION NUMBER: 60/088876  
 ; PRIOR FILING DATE: 1998-06-11  
 ; PRIOR APPLICATION NUMBER: 60/089105  
 ; PRIOR FILING DATE: 1998-06-12  
 ; PRIOR APPLICATION NUMBER: 60/089440  
 ; PRIOR FILING DATE: 1998-06-16  
 ; PRIOR APPLICATION NUMBER: 60/089512  
 ; PRIOR FILING DATE: 1998-06-16  
 ; PRIOR APPLICATION NUMBER: 60/089514  
 ; PRIOR FILING DATE: 1998-06-16  
 ; PRIOR APPLICATION NUMBER: 60/089532  
 ; PRIOR FILING DATE: 1998-06-17  
 ; PRIOR APPLICATION NUMBER: 60/089538  
 ; PRIOR FILING DATE: 1998-06-17  
 ; PRIOR APPLICATION NUMBER: 60/089598  
 ; PRIOR FILING DATE: 1998-06-17  
 ; PRIOR APPLICATION NUMBER: 60/089599  
 ; PRIOR FILING DATE: 1998-06-17  
 ; PRIOR APPLICATION NUMBER: 60/089600  
 ; PRIOR FILING DATE: 1998-06-17  
 ; PRIOR APPLICATION NUMBER: 60/089653  
 ; PRIOR FILING DATE: 1998-06-17  
 ; PRIOR APPLICATION NUMBER: 60/089801  
 ; PRIOR FILING DATE: 1998-06-18  
 ; PRIOR APPLICATION NUMBER: 60/089907  
 ; PRIOR FILING DATE: 1998-06-18  
 ; PRIOR APPLICATION NUMBER: 60/089908  
 ; PRIOR FILING DATE: 1998-06-18  
 ; PRIOR APPLICATION NUMBER: 60/089947  
 ; PRIOR FILING DATE: 1998-06-19  
 ; PRIOR APPLICATION NUMBER: 60/089948  
 ; PRIOR FILING DATE: 1998-06-19  
 ; PRIOR APPLICATION NUMBER: 60/089952  
 ; PRIOR FILING DATE: 1998-06-19  
 ; PRIOR APPLICATION NUMBER: 60/090246  
 ; PRIOR FILING DATE: 1998-06-22  
 ; PRIOR APPLICATION NUMBER: 60/090255  
 ; PRIOR FILING DATE: 1998-06-22  
 ; PRIOR APPLICATION NUMBER: 60/090254  
 ; PRIOR FILING DATE: 1998-06-22  
 ; PRIOR APPLICATION NUMBER: 60/090349  
 ; PRIOR FILING DATE: 1998-06-23  
 ; PRIOR APPLICATION NUMBER: 60/090355  
 ; PRIOR FILING DATE: 1998-06-23  
 ; PRIOR APPLICATION NUMBER: 60/090429  
 ; PRIOR FILING DATE: 1998-06-24  
 ; PRIOR APPLICATION NUMBER: 60/090431  
 ; PRIOR FILING DATE: 1998-06-24  
 ; PRIOR APPLICATION NUMBER: 60/090435  
 ; PRIOR FILING DATE: 1998-06-24  
 ; PRIOR APPLICATION NUMBER: 60/090444  
 ; PRIOR FILING DATE: 1998-06-24  
 ; PRIOR APPLICATION NUMBER: 60/090535  
 ; PRIOR FILING DATE: 1998-06-24  
 ; PRIOR APPLICATION NUMBER: 60/090540  
 ; PRIOR FILING DATE: 1998-06-24  
 ; PRIOR APPLICATION NUMBER: 60/090542  
 ; PRIOR FILING DATE: 1998-06-24  
 ; PRIOR APPLICATION NUMBER: 60/090557  
 ; PRIOR FILING DATE: 1998-06-24



Sequence 2976, Application US/09134001C  
 Patent No. 6380370  
 GENERAL INFORMATION:  
 APPLICANT: Lynn Doucette-Stamm et al.  
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
 TITLE OF INVENTION: EPIDERMIDS FOR DIAGNOSTICS AND THERAPEUTICS  
 FILE REFERENCE: GTC-007  
 CURRENT APPLICATION NUMBER: US/09/134, 001C  
 CURRENT FILING DATE: 1998-08-13  
 PRIOR APPLICATION NUMBER: US 60/064, 964  
 PRIOR FILING DATE: 1997-11-08  
 PRIOR APPLICATION NUMBER: US 60/055, 779  
 PRIOR FILING DATE: 1997-08-14  
 NUMBER OF SEQ ID NOS: 5674  
 SEQ ID NO: 2976  
 LENGTH: 296  
 TYPE: PRT  
 ORGANISM: Staphylococcus epidermidis

US-09-134-001C-2976

Query Match 4.8%; Score 87; DB 4; Length 431;  
 Best Local Similarity 22.2%; Pred. No. 1.4;  
 Matches 51; Conservative 49; Mismatches 98; Indels 32; Gaps 14;

Qy            124 HFF--DNDENYGKG-IEWTRKKNMFPSYQQTIEKSPAYFITERVPIRYYKOMSSIKLII 180  
 Db            152 HFADDEGQSFSSETFIQLIKQRKD--PARIETLGDSRTEFLFPELFREEQ---IL 204  
 Qy            181 IVREPTTRAI-SDYTQVLEGERKRNKTYKFEKLADENTCE--VNTKYKAVRTSIV 234  
 Db            205 PIKEKNYLAILRLSCNQITEKVRVERIEETDYRQLTIVTEKLDLSFLWLNSEFESNKSKL 264  
 Qy            235 TKHL-EWLKYFPLEQFHVDGDBLITEPLPELO-LVEKFLNLPRISQVNLYNATRGF 292  
 Db            265 TKVLQEWIRY--VIGH---DRLVEEEERYQETLEXPKNSDKDLYQERLAQQALAH 318  
 Qy            293 YCLRFNLLFKLAGSKG--RIHPEVDPSVITYKURKFHHFENOKFYQITG 340  
 Db            319 PAVERGMNLKLTQ-ERGYLDVH--PEV--KIVFFFYEQEARLVG 362

RESULT 13  
 US-08-332-562A-136  
 Sequence 136, Application US/08332562A  
 Patent No. 5985539  
 GENERAL INFORMATION:  
 APPLICANT: MCKENZIE, Ian F.C.  
 TITLE OF INVENTION: NUCLEAR ACID AND AMINO ACID SEQUENCES RELATING TO  
 ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS  
 NUMBER OF SEQUENCES: 7310  
 CORRESPONDENCE ADDRESS:  
 GENOME THERAPEUTICS CORPORATION  
 STREET: 100 Beaver Street  
 CITY: Waltham  
 STATE: Massachusetts  
 COUNTRY: USA  
 ZIP: 02354  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: CD-ROM ISO9660  
 COMPUTER: PC  
 OPERATING SYSTEM: <Unknown>  
 SOFTWARE: ASCII  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/107, 532A  
 FILING DATE: 30-Jun-1998  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 60/085, 598  
 FILING DATE: 14 May 1998  
 APPLICATION NUMBER: 60/051571  
 FILING DATE: July 2, 1997  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Arianiello, Pamela Deneke  
 REGISTRATION NUMBER: 40,489  
 REQUENCE/DOCKET NUMBER: GTC-012  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (781) 893-5007  
 TELEFAX: (781) 893-8277

RESULT 12  
 US-09-107, 532A-5796  
 Sequence 5796, Application US/09107532A  
 Patent No. 6583275  
 GENERAL INFORMATION:  
 APPLICANT: Lynn A. Doucette-Stamm and David Bush  
 TITLE OF INVENTION: NUCLEAR ACID AND AMINO ACID SEQUENCES RELATING TO  
 ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS  
 NUMBER OF SEQUENCES: 7310  
 CORRESPONDENCE ADDRESS:  
 GENOME THERAPEUTICS CORPORATION  
 STREET: 100 Beaver Street  
 CITY: Waltham  
 STATE: Massachusetts  
 COUNTRY: USA  
 ZIP: 02354  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: CD-ROM ISO9660  
 COMPUTER: PC  
 OPERATING SYSTEM: <Unknown>  
 SOFTWARE: ASCII  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/107, 532A  
 FILING DATE: 30-Jun-1998  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 60/085, 598  
 FILING DATE: 14 May 1998  
 APPLICATION NUMBER: 60/051571  
 FILING DATE: July 2, 1997  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Arianiello, Pamela Deneke  
 REGISTRATION NUMBER: 40,489  
 REQUENCE/DOCKET NUMBER: GTC-012  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (781) 893-5007  
 TELEFAX: (781) 893-8277

RESULT 13  
 US-08-332-562A-136  
 Sequence 136, Application US/08332562A  
 Patent No. 5985539  
 GENERAL INFORMATION:  
 APPLICANT: MCKENZIE, Ian F.C.  
 TITLE OF INVENTION: NUCLEAR ACID AND AMINO ACID SEQUENCES RELATING TO  
 ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS  
 NUMBER OF SEQUENCES: 7310  
 CORRESPONDENCE ADDRESS:  
 GENOME THERAPEUTICS CORPORATION  
 STREET: 100 Beaver Street  
 CITY: Waltham  
 STATE: Massachusetts  
 COUNTRY: USA  
 ZIP: 02354  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.3.0  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/332, 562A  
 FILING DATE: 31-Oct-1994  
 CLASSIFICATION: 435  
 PRIOR APPLICATION NUMBER: US 07/896, 457  
 APPLICATION NUMBER: US 07/896, 457  
 FILING DATE: 27-May-1992  
 ATTORNEY/AGENT INFORMATION:  
 NAME: BENT, Stephen A.  
 REGISTRATION NUMBER: 29,768  
 REQUENCE/DOCKET NUMBER: 54270/119/GRHA  
 TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 672-5300  
 TELEFAX: (202) 672-5399  
 TELEX: 904136  
 INFORMATION FOR SEQ ID NO: 136:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 283 amino acids  
 TYPE: amino acid  
 TOPOLOGY: lineair  
 MOLECULE TYPE: protein  
 US-08-332-562A-136

Query Match Score 4.8%; DB 2; Length 283;  
 Best Local Similarity 23.9%; Pred. No. 0.84; Gaps 10;  
 Matches 53; Conservative 31; Mismatches 85; Indels 53; Gaps 10;

Qy 45 GRIGGARTQAEFPFLRQLQFKRLHHEFRKGNAKSEQVRLHD----LVQQLPKAI 94  
 Db 74 GRISIRSQVQASITFKATVNDSG---EYR---CQEQTRLSDPVDLGVTSWILQLQVL 127  
 Qy 95 IIGVRKGCTGTRALLEMLNHPAVKKAQSEIHFNDENYGGKIEWRKCKMPSSYPOQ--- 150  
 Db 128 FL---EGETIT---LRCHSWENKLNRISPFNE---KSTRYHYSNSESPICKANHSH 176  
 Qy 151 -----TTEKSFAVFTTEEVPERIYKGNSSIKLLIVREPTTRAISDYTOVLEG 199  
 Db 177 SGDYYCKGSGLRTLHQSKPVTIVQGP---KSSRSLEPLTVTAATGIAVIL-- 230

Qy 200 KERKNCYKTYKFERKLAIDPNTCEVNTKYKAVRTSIYT--KHLK 239  
 Db 231 --VSLLVKKQVPDNPPDLEBAAKTEAMITYSLKHPE 269

RESULT 14 US-09-134-001C-3346  
 ; Sequence 3346, Application US/09134001C  
 ; General Information:  
 ; Applicant: Lynn Doucette-Stamm et al.  
 ; Title of Invention: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
 ; File Reference: GTC-007  
 ; Current Application Number: US/09/134, 001C  
 ; Current Filing Date: 1998-08-13  
 ; Prior Application Number: US 60/064, 964  
 ; Prior Filing Date: 1997-11-08  
 ; Prior Application Number: US 60/055, 779  
 ; Prior Filing Date: 1997-08-14  
 ; SEQ ID NO: 5674  
 ; SEQ ID NO: 5674  
 ; SEQ ID NO: 4200  
 ; TYPE: PRT  
 ; ORGANISM: staphylococcus epidermidis

Query Match Score 4.8%; DB 4; Length 467;  
 Best Local Similarity 22.9%; Pred. No. 1.8;  
 Matches 52; Conservative 31; Mismatches 73; Indels 71; Gaps 13;

Qy 52 TQAEF PLRAIQFKRLHHEFRKGNAKSEQVRLHDIVQ-----LPKALIIGVRKG-G 102  
 Db 269 SQAQTIAILDMSMNGNIKHPF----ETVKKAESTONQAEDSILKPFTLVEKGEG 321

Qy 103 TRA---LLEMNLHPAVVKASQEIHFNDENYGR-----GIEYRKKMPFSYPQQITI 153  
 Db 322 LKAEEFLQQTINANPITSGPD--FTEDMLGYCLPNESGALYREG----- 367

Qy 154 EKSPAYFITEVPERIYKGNSSIKLLIVREPTTRAISDYTOVLEGKE----- 201  
 Db 368 ----FITHLMP---KDKSYLVIDNFNRIDVDFQTYINVLEYEVTLPRYNKDGS 417

Qy 202 --RKNT-TYKFKELAIDPMTCEVNTKYKAVRTSITYKLERWYK 244  
 Db 418 IKWSRQKDSPFH---NPNWHIVGITYDSIE-KIKQKYSQSOFIY 458

SEQ ID NO: 3346  
 LENGTH: 426  
 TYPE: PRT  
 ORGANISM: staphylococcus epidermidis

US-09-134-001C-3346

Query Match Score 4.8%; DB 4; Length 426;  
 Best Local Similarity 21.9%; Pred. No. 1.6; Gaps 14;  
 Matches 64; Conservative 38; Mismatches 89; Indels 101; Gaps 14;

Qy 59 RALQFKRLHHEFRKGNAKSEQVRLHD-----VQPKALIIGVRKGTRALEM 109  
 Db 8 KALTFWQI1----SNEBRKVMTHTITIVGIGNYGIDELPLGTRFLER-E SKVYART 61

Qy 110 IMLHPAVVKASQEIHFNDENYGGKEMWYKMPFSYPQQITIERSPAYFTEVERI 169  
 Db 62 LN-HPVINTLKCETEFESFDST-----EAHDFEDYEA 96

Qy 170 YRNNSSIKL---I-----VREPTTRAISDY-----QVLEGGERKNTKYYK 209  
 Db 97 --VTSLLFELAQSEDIVYAVGHPRVAETTVKLEYHNRKDISVYLGGSFID 151

Qy 210 FEKLAIDN----TCEVNTKYKAVRTSITYKHLERWLYKFPIEQ---FHV- 252  
 Db 152 FEAVDVENDGFTFLDGTSLKESALNVRTNTVIQVSVMAADLKLTLMERYPPDDENVK 211

Search completed: December 4, 2003, 16:46:41  
 Job time : 31 secs